# STIC Search Report
## Biotech-Chem Library

## STIC Database Tracking Number: 142881

**TO: Jeffrey Parkin**
**Location: REM 3d39 / 3c18**
**Art Unit: 1648**
**Tuesday, January 25, 2005**

**Case Serial Number: 09/319156**

**From: Deirdre Arnold**
**Location: Biotech-Chem Library**
**REM 1A64**
**Phone: 571-272-2532**

**Deirdre.Arnold@uspto.gov**

## Search Notes

**Please contact me if you need to search the pending files.**

Please contact me if you encounter any problems with the file or the disk.

*Please feel free to contact me if you have any questions or would like to amend the search.*

Thank you for using STIC services.

Regards,
Deirdre Arnold

~~142354~~ 142881

**STIC-Biotech/ChemLib**

**From:** Parkin, Jeffrey
**Sent:** Tuesday, January 11, 2005 8:50 PM
**To:** STIC-Biotech/ChemLib
**Subject:** SEQUENCE SEARCH: U.S. Serial No. 09/319,156

Please search **SEQ ID NOS.: 6, 9, 10**, and **12** from the aforementioned application (U.S. Serial No. **09/319,156**) v. all relevant databases.

Please provide the first **30 SUMMARIES**.

Place results on both paper and disk.

Thanks!

JSP
AU 1648
REM-3D39
2-0908

<http://expoweb1:8001/cgi-bin/expo/GenInfo/snquery.pl?APPL_ID=09319156>.

*************************

STAFF USE ONLY

Searcher: Arnold
Searcher Phone: 2- 2532
Date Searcher Picked up: 1/19/05
Date Completed: 1/25/05
Searcher Prep/Rev. Time:___
Online Time:____________

*************************

Type of Search
NA Sequence: # 3
AA Sequence :# 1
Structure: #____________
Bibliographic:____________
Litigation:____________
Patent Family:____________
Other:____________

*************************

Vendors and cost where applicable
STN:____________
DIALOG:____________
QUESTEL/ORBIT:____________
LEXIS/NEXIS:____________
SEQUENCE SYSTEM:________
WWW/Internet:____________
Other(Specify):____________

```
                     GenCore version 5.1.6
             Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 10:09:14 ; Search time 2663.68 Seconds
                                           (without alignments)
                                           11273.477 Million cell updates/sec

Title:          US-09-319-156B-6
Perfect score:  635
Sequence:       1 ccctgtatctttaacctcct..........tgaaaaaaaaaaaaaaaaa 635

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters:      9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                      %
Result              Query
  No.     Score     Match Length DB  ID                        Description
-------------------------------------------------------------------------------
    1       635     100.0    635  6  AR344385                  AR344385 Sequence
    2       635     100.0    635  6  AX001024                  AX001024 Sequence
    3       635     100.0    635  6  BD136195                  BD136195 Retrovira
    4       635     100.0   2004 14  AF127229                   AF127229 Multiple
c   5     631.8      99.5 167694  9  AC113137                  AC113137 Homo sapi
c   6     631.8      99.5 176095  2  AC019346                  AC019346 Homo sapi
    7     631.8      99.5 210336  2  AC022171                  AC022171 Homo sapi
c   8     580.2      91.4 148724  2  AC010778                  AC010778 Homo sapi
    9     579.2      91.2   8339  9  AL162912                  AL162912 Human DNA
c  10       571      89.9 136901  9  AC073626                  AC073626 Homo sapi
c  11     566.4      89.2 176188  9  AC068898                  AC068898 Homo sapi
   12     564.6      88.9  89728  9  AL583805                  AL583805 Human DNA
   13     560.2      88.2 112405  9  AC072023                  AC072023 Homo sapi
c  14     559.8      88.2 179114  9  BS000045                  BS000045 Pan trogl
c  15     556.6      87.7 174019  9  AP001538                  AP001538 Homo sapi
c  16     556.6      87.7 340000  9  AP001674                  AP001674 Homo sapi
   17     551.8      86.9 149755  2  AP001545                  AP001545 Homo sapi
c  18     551.8      86.9 152980  2  AC090313                  AC090313 Homo sapi
c  19     551.8      86.9 163803  9  AC093531                  AC093531 Homo sapi
c  20     551.8      86.9 176425  9  AC107075                  AC107075 Homo sapi
c  21     551.4      86.8 128468  9  AL139090                  AL139090 Human DNA
   22     551.4      86.8 140756  9  AL139038                  AL139038 Human DNA
c  23     551.4      86.8 167366  2  AC021774                  AC021774 Homo sapi
   24     551.4      86.8 183499  2  AL607153                  AL607153 Homo sapi
c  25     550.2      86.6 164211  2  AP002790                  AP002790 Homo sapi
   26     550.2      86.6 167843  9  AC064801                  AC064801 Homo sapi
   27       549      86.5 180523  9  CNS01DSH                  AL121784 Human chr
   28       549      86.5 190565  9  AC007374                  AC007374 Homo sapi
c  29       549      86.5 203777  9  CNS01DVH                  AL135818 Human chr
   30     547.8      86.3 112544  9  AC109992                  AC109992 Homo sapi
   31     547.6      86.2 192178  2  AC009443                  AC009443 Homo sapi
   32     547.6      86.2 192899  9  CNS05TF5                  AL365295 Human chr
   33     547.4      86.2 172281  9  AC068492                  AC068492 Homo sapi
c  34     547.2      86.2 169418  9  AC104163                  AC104163 Homo sapi
   35     546.6      86.1 161054  9  AC104444                  AC104444 Homo sapi
   36     546.4      86.0   2074  6  AX478550                  AX478550 Sequence
c  37       546      86.0 169234  9  CNS06C7Q                  AL390799 Human chr
   38       545      85.8   1329  6  AR344389                  AR344389 Sequence
   39       545      85.8   1329  6  AX001030                  AX001030 Sequence
   40       545      85.8   1329  6  BD136199                  BD136199 Retrovira
   41       545      85.8  99408  9  AL359385                  AL359385 Human DNA
   42       545      85.8 170586  2  AL356126                  AL356126 Homo sapi
c  43     543.8      85.6 181166  9  AC087897                  AC087897 Homo sapi
c  44     543.4      85.6  88328  9  AL357874                  AL357874 Human DNA
   45       539      84.9   2082  2  AC134379                  AC134379 Homo sapi


                              ALIGNMENTS


RESULT 1
AR344385
LOCUS       AR344385                 635 bp    mRNA    linear   PAT 17-AUG-2003
DEFINITION  Sequence 102 from patent US 6582703.
ACCESSION   AR344385
VERSION     AR344385.1  GI:33740326
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Perron,H., Beseme,F., Bedin,F., Paranhos-Baccala,G.,
            Komurian-Pradel,F., Jolivet-Reynaud,C. and Mandrand,B.
  TITLE     Isolated nucleotide sequences associated with multiple sclerosis or
            rheumatoid arthritis and a process of detecting
  JOURNAL   Patent: US 6582703-A 102 24-JUN-2003;
FEATURES             Location/Qualifiers
     source          1..635
                     /organism="unknown"
                     /mol_type="mRNA"
ORIGIN

  Query Match             100.0%;  Score 635;  DB 6;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-189;
  Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
```

```
Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
```

```
RESULT 2
AX001024
LOCUS       AX001024                 635 bp    DNA     linear   PAT 10-MAR-2000
DEFINITION  Sequence 6 from Patent WO9902666.
ACCESSION   AX001024
VERSION     AX001024.1  GI:7241262
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Ott,C. and Bedin,F.
  TITLE     RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR
            ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR
            DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
  JOURNAL   Patent: WO 9902666-A 6 21-JAN-1999;
            BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 635;  DB 6;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-189;
  Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
```

```
RESULT 3
BD136195
LOCUS       BD136195                 635 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Retroviral nucleic material and nucleotide fragments, in
            particular, associated with multiple sclerosis and/or rheumatoid
            arthritis, for diagnostic, prophylactic and therapeutic uses.
ACCESSION   BD136195
VERSION     BD136195.1  GI:23231140
KEYWORDS    JP 2002509437-A/5.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Baccala,G.P., Pradel,F.K., Bedin,F., Sodoyer,M., Ott,C., Mallet,F.,
            Perron,H. and Mandrand,B.
  TITLE     Retroviral nucleic material and nucleotide fragments, in
            particular, associated with multiple sclerosis and/or rheumatoid
            arthritis, for diagnostic, prophylactic and therapeutic uses
  JOURNAL   Patent: JP 2002509437-A 5 26-MAR-2002;
            BIO MERIEUX
COMMENT     OS   Unidentified
            PN   JP 2002509437-A/5
            PD   26-MAR-2002
            PF   07-JUL-1998 JP 1999508255
            PR   07-JUL-1997 FR      97/08816
            PI   GLAUCIA PARAHNOS BACCALA,FLORENCE KOMURIAN PRADEL,FREDERIC  PI
              BEDIN,
            PI   MIREILLE SODOYER,CATHERINE OTT,FRANCOIS MALLET,HERVE PERRON,
            PI   BERNARD MANDRAND
            PC   C12N15/48,C12Q1/70,C07K14/15,A61K31/70
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   Retroviral nucleic material and nucleotide fragments, in  CC
                      particular,
            CC   associated with multiple sclerosis and/or
            rheumatoid arthritis,
            CC             for
            CC   diagnostic, prophylactic and therapeutic uses FH   Key
                Location/Qualifiers
            FT   source             1. .635
            FT                      /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 635;  DB 6;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-189;
  Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635


RESULT 4
AF127229
LOCUS       AF127229                2004 bp    mRNA    linear   VRL 11-AUG-1999
DEFINITION  Multiple sclerosis associated retrovirus element clone CL6
            pol-env/3'LTR-like mRNA sequence.
ACCESSION   AF127229
VERSION     AF127229.1  GI:5726294
KEYWORDS    .
SOURCE      Multiple sclerosis associated retrovirus element
  ORGANISM  Multiple sclerosis associated retrovirus element
            Viruses; Retroid viruses; Retroviridae.
REFERENCE   1  (bases 1 to 2004)
  AUTHORS   Komurian-Pradel,F., Paranhos-Baccala,G., Bedin,F.,
            Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,
            Mallet,F., Mandrand,B. and Perron,H.
  TITLE     Molecular cloning and characterization of MSRV-related sequences
            associated with retrovirus-like particles
  JOURNAL   Virology 260 (1), 1-9 (1999)
  MEDLINE   99335590
   PUBMED   10405350
```

```
REFERENCE   2  (bases 1 to 2004)
  AUTHORS   Komurian-Pradel,F., Paranhos-Baccala,G., Bedin,F.,
            Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,
            Mallet,F., Mandrand,B. and Perron,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1999) UMR103 CNRS, bioMerieux, 46, allee
            D'Italie, Lyon 69007, France
FEATURES             Location/Qualifiers
     source          1..2004
                     /organism="Multiple sclerosis associated retrovirus
                     element"
                     /mol_type="mRNA"
                     /db_xref="taxon:89382"
                     /clone="CL6"
     misc_feature    1..2004
                     /note="similar to pol-env and 3'LTR region"
ORIGIN

  Query Match             100.0%;  Score 635;  DB 14;  Length 2004;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-189;
  Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1370 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 1429

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1430 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 1489

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1490 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 1549

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1550 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 1609

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1610 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 1669

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1670 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 1729

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1730 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 1789

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1790 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 1849

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1850 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 1909

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1910 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 1969

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db       1970 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 2004


RESULT 5
AC113137/c
LOCUS       AC113137               167694 bp    DNA     linear   PRI 25-APR-2002
DEFINITION  Homo sapiens chromosome 18, clone RP11-622J9, complete sequence.
ACCESSION   AC113137
VERSION     AC113137.2  GI:19683504
```

```
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167694)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-622J9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167694)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 167694)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 167694)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 22, 2002 this sequence version replaced gi:18875226.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L24497
                Center clone name: 622_J_9
            --------------.
FEATURES             Location/Qualifiers
     source          1. .167694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-622J9"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   2566. .2593
                     /rpt_family="AT_rich"
     repeat_region   3108. .3127
                     /rpt_family="(TA)n"
     repeat_region   3128. .3153
                     /rpt_family="(TG)n"
     repeat_region   3154. .3208
                     /rpt_family="(TA)n"
     repeat_region   complement(3212. .3818)
                     /rpt_family="L1MA9"
     repeat_region   4840. .5079
                     /rpt_family="L1ME3"
     repeat_region   5234. .6028
                     /rpt_family="L1ME3"
     repeat_region   complement(6259. .6470)
                     /rpt_family="MIR"
     repeat_region   6723. .6750
                     /rpt_family="AT_rich"
     repeat_region   6982. .7280
                     /rpt_family="AluY"
     repeat_region   7780. .7811
                     /rpt_family="AT_rich"
     repeat_region   complement(8120. .8302)
                     /rpt_family="L1M4"
     repeat_region   complement(8444. .8971)
                     /rpt_family="L1ME3A"
     repeat_region   complement(8973. .9386)
                     /rpt_family="L1ME3A"
     repeat_region   complement(9387. .9690)
                     /rpt_family="L1MA9"
     repeat_region   9898. .10497
                     /rpt_family="MLT1E3"
     repeat_region   11218. .11555
                     /rpt_family="L2"
     repeat_region   complement(12442. .13393)
                     /rpt_family="L1PA16"
     repeat_region   13550. .13907
```

```
                     /rpt_family="L1ME"
     repeat_region   14238..15632
                     /rpt_family="L1PB1"
     repeat_region   15633..15912
                     /rpt_family="(TA)n"
     repeat_region   15920..16279
                     /rpt_family="L1PB1"
     repeat_region   16338..17100
                     /rpt_family="HAL1b"
     repeat_region   17572..17628
                     /rpt_family="L1M4"
     repeat_region   17684..18430
                     /rpt_family="L1PA4"
     repeat_region   18441..19292
                     /rpt_family="L2"
     repeat_region   19293..20623
                     /rpt_family="L1MB3"
     repeat_region   complement(20633..22006)
                     /rpt_family="L1P"
     repeat_region   22008..22831
                     /rpt_family="L1MB3"
     repeat_region   22853..23516
                     /rpt_family="L2"
     repeat_region   24420..24468
                     /rpt_family="AT_rich"
     repeat_region   24886..24913
                     /rpt_family="AT_rich"
     unsure          25154..25156
                     /note="single clone coverage"
     unsure          25189..25193
                     /note="<30 qual SNGL region"
     repeat_region   complement(26144..26248)
                     /rpt_family="MIR"
     repeat_region   27732..27794
                     /rpt_family="AT_rich"
     repeat_region   27889..27937
                     /rpt_family="(TA)n"
     repeat_region   28013..28078
                     /rpt_family="(TAGA)n"
     repeat_region   28704..28743
                     /rpt_family="(TAGA)n"
     repeat_region   29367..29451
                     /rpt_family="GA-rich"
     repeat_region   complement(30946..31096)
                     /rpt_family="MLT1H"
     repeat_region   31912..31989
                     /rpt_family="MER67C"
     repeat_region   32017..32303
                     /rpt_family="MER67C"
     repeat_region   complement(32304..34133)

  Query Match             99.5%;  Score 631.8;  DB 9;  Length 167694;
  Best Local Similarity   99.7%;  Pred. No. 2.9e-188;
  Matches  633;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     123609 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACCGTAAAACTA 123550

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123549 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 123490

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123489 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 123430

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123429 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 123370

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     123369 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 123310

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123309 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 123250

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123249 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 123190

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123189 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 123130

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123129 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 123070

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123069 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 123010

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||| ||||
Db     123009 ATTAAATCTTGCAACTGAAAAAAAAAAAAAGAAAA 122975


RESULT 6
AC019346/c
LOCUS       AC019346               176095 bp    DNA     linear   HTG 25-OCT-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-497M7 map 18, WORKING DRAFT
            SEQUENCE.
ACCESSION   AC019346
VERSION     AC019346.4  GI:16418201
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176095)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-497M7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176095)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 25, 2001 this sequence version replaced gi:7230183.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
```

```
          -------------- Project Information
              Center project name: L4296
              Center clone name: 497_M_7
          -------------- Summary Statistics
              Sequencing vector: M13; M77815; 32% of reads
              Sequencing vector: Plasmid; n/a; 68% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 176087 bases at least Q40
              Consensus quality: 176095 bases at least Q30
              Consensus quality: 176095 bases at least Q20
              Insert size: 178000; agarose-fp
              Insert size: 176095; sum-of-contigs
              Quality coverage: 11.5 in Q20 bases; agarose-fp
              Quality coverage: 11.7 in Q20.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1  176095: contig of 176095 bp in length.
          assembly_fragment.
FEATURES             Location/Qualifiers
     source          1..176095
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-497M7"
                     /clone_lib="RPCI-11 Human Male BAC"
ORIGIN

  Query Match             99.5%;  Score 631.8;  DB 2;  Length 176095;
  Best Local Similarity   99.7%;  Pred. No. 2.9e-188;
  Matches  633;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     150524 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACCGTAAAACTA 150465

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150464 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 150405

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150404 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 150345

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150344 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 150285

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150284 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 150225

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150224 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 150165

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150164 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 150105

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150104 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 150045
```

```
Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150044 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 149985

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     149984 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 149925

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||| ||||
Db     149924 ATTAAATCTTGCAACTGAAAAAAAAAAAAAGAAAA 149890


RESULT 7
AC022171
LOCUS       AC022171              210336 bp    DNA     linear   HTG 26-MAR-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-407C18, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC022171
VERSION     AC022171.18  GI:13270574
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 210336)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 210336)
  AUTHORS   Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
            Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
            Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
            Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 10, 2001 this sequence version replaced gi:13122770.
            -------------- Genome Center
                Center: Stanford DNA Sequencing and Technology Development
            Center
                Center code: SDSTDC
                Web site: http://sequence-www.stanford.edu/group/human/
                Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                Center project name: 758
                Center clone name: RP11-407C18
            -------------- Summary Statistics
                Sequencing Vector: M13mp18; X02513; 97% of reads
                Sequencing Vector: plasmid; plasmid_accession; 3% of reads
                Chemistry: Dye-primer; 10% of reads
                Chemistry: Dye-terminator Big Dye; 88% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 208312 bases at least Q40
                Consensus quality: 208877 bases at least Q30
                Consensus quality: 209157 bases at least Q20
                Insert size: 199762; agarose-fp
                Insert size: 210236; sum-of-contigs
                Quality coverage:  8.7x in Q20 bases; agarose-fp
                Quality coverage:  8.3x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     66468: contig of 66468 bp in length
            *    66469    66568: gap of unknown length
            *    66569   210336: contig of 143768 bp in length.
FEATURES             Location/Qualifiers
     source          1..210336
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /clone="RP11-407C18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1..66468
                     /note="assembly_name:Contig44
                     clone_end:T7"
     misc_feature    66569..210336
                     /note="assembly_name:Contig45
                     clone_end:SP6"
ORIGIN

  Query Match             99.5%;  Score 631.8;  DB 2;  Length 210336;
  Best Local Similarity   99.7%;  Pred. No. 3e-188;
  Matches  633;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      58176 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACCGTAAAACTA 58235

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58236 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 58295

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58296 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 58355

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58356 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 58415

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58416 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 58475

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58476 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 58535

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58536 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 58595

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58596 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 58655

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58656 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 58715

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58716 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 58775

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||| ||||
Db      58776 ATTAAATCTTGCAACTGAAAAAAAAAAAAAGAAAA 58810


RESULT 8
AC010778/c
LOCUS       AC010778              148724 bp    DNA     linear   HTG 30-MAR-2000
DEFINITION  Homo sapiens clone RP11-2N15, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC010778
```

```
VERSION     AC010778.3  GI:7341824
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148724)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-2N15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148724)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2000 this sequence version replaced gi:6087973.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2677
                Center clone name: 2_N_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 140899 bases at least Q40
                Consensus quality: 144351 bases at least Q30
                Consensus quality: 145949 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 148024; sum-of-contigs
                Quality coverage: 3.9 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2462: contig of 2462 bp in length
            *     2463     2562: gap of 100 bp
            *     2563     7948: contig of 5386 bp in length
            *     7949     8048: gap of 100 bp
            *     8049    15800: contig of 7752 bp in length
            *    15801    15900: gap of 100 bp
            *    15901    25059: contig of 9159 bp in length
            *    25060    25159: gap of 100 bp
            *    25160    41906: contig of 16747 bp in length
            *    41907    42006: gap of 100 bp
            *    42007    63341: contig of 21335 bp in length
            *    63342    63441: gap of 100 bp
            *    63442    98587: contig of 35146 bp in length
```

```
           *     98588    98687: gap of 100 bp
           *     98688   148724: contig of 50037 bp in length.
FEATURES             Location/Qualifiers
     source          1..148724
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-2N15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..2462
                     /note="assembly_fragment"
     misc_feature    2563..7948
                     /note="assembly_fragment"
     misc_feature    8049..15800
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    15901..25059
                     /note="assembly_fragment"
     misc_feature    25160..41906
                     /note="assembly_fragment"
     misc_feature    42007..63341
                     /note="assembly_fragment"
     misc_feature    63442..98587
                     /note="assembly_fragment"
     misc_feature    98688..148724
                     /note="assembly_fragment"
ORIGIN

  Query Match             91.4%;  Score 580.2;  DB 2;  Length 148724;
  Best Local Similarity   94.8%;  Pred. No. 6.7e-172;
  Matches  600;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db      40271 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAACTGTAAAACTA 40212

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||||||||| |||||| ||| ||
Db      40211 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGACCCACCGCGGATCC 40152

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||| |||||||||||||||||||  |||||||| ||| ||||||||||||||||||
Db      40151 CTGGACCAGCCTGCTAGCCCATGCTCCAGTGTTAATGGCATCGAAGGCACCCCTCCCGAG 40092

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||| ||
Db      40091 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 40032

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||
Db      40031 CATCAGCCAACTTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAGAGAC 39972

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||| 
Db      39971 AGGACTAGCTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 39912

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               || |||||| |||||||||||||||||||||||||||||||| ||||||||||||||||
Db      39911 ACACCCACCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCAACCAATCAGAGAGCTC 39852

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      39851 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 39792

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| ||||||||||||||||||||||||||||||||||||  ||| ||||| |
Db      39791 AGAGCACAGCTGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCTGGCCAGCAACAG 39732

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      39731 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 39672

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAA 633
              ||||||||||||||||||||||||||||   ||
Db      39671 ATTAAATCTTGCAACTGAAAAAAAAAAAATGCAA 39639


RESULT 9
AL162912
LOCUS       AL162912                8339 bp    DNA     linear   PRI 22-NOV-2001
DEFINITION  Human DNA sequence from clone RP6-198C4 on chromosome Xq13.3-21.2,
            complete sequence.
ACCESSION   AL162912 AL121824
VERSION     AL162912.1  GI:7406722
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 8339)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP6-198C4 is from the
            library RPCI-6 constructed by the group of Pieter de Jong. For
            further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pPAC4
            IMPORTANT: This sequence is not the entire insert of clone
            RP6-198C4 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
FEATURES             Location/Qualifiers
     source          1..8339
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q13.3-21.2"
                     /clone="RP6-198C4"
                     /clone_lib="RPCI-6"
     repeat_region   1..5641
                     /note="HERV17 repeat: matches 368..5930 of consensus"
     repeat_region   5635..5673
                     /note="HERV17 repeat: matches 7419..7457 of consensus"
     repeat_region   5664..5955
                     /note="HERV17 repeat: matches 8244..8523 of consensus"
     repeat_region   5956..6136
                     /note="LTR17 repeat: matches 1..239 of consensus"
     repeat_region   6195..6280
                     /note="LTR17 repeat: matches 239..324 of consensus"
ORIGIN
```

```
  Query Match             91.2%;  Score 579.2;  DB 9;  Length 8339;
  Best Local Similarity   94.8%;  Pred. No. 8.4e-172;
  Matches  599;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db       5664 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAACTGTAAAACTA 5723

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| |||||||||||||||| |||||| ||| ||
Db       5724 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGACCCACCGCGGATCC 5783

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||| |||||||||||||||||||  |||||||| ||| ||||||||||||||||||
Db       5784 CTGGACCAGCCTGCTAGCCCATGCTCCAGTGTTAATGGCATCGAAGGCACCCCTCCCGAG 5843

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||| ||
Db       5844 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 5903

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||
Db       5904 CATCAGCCAACTTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAGAGAC 5963

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       5964 AGGACTAGCTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 6023

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               || |||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       6024 ACACCCACCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCAACCAATCAGAGAGCTC 6083

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       6084 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 6143

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| ||||||||||||||||||||||||||||||||||||  |||| ||||| |
Db       6144 AGAGCACAGCTGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCTGGCCAGCAACAG 6203

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       6204 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 6263

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAA 632
              |||||||||||||||||||||||||||   ||
Db       6264 ATTAAATCTTGCAACTGAAAAAAAAAATGCAA 6295


RESULT 10
AC073626/c
LOCUS       AC073626              136901 bp    DNA     linear   PRI 30-JAN-2004
DEFINITION  Homo sapiens BAC clone RP11-95P9 from 7, complete sequence.
ACCESSION   AC073626
VERSION     AC073626.7  GI:12863221
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 136901)
  AUTHORS   Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
            Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
            Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
            Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
            Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
            Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L.,
            Kalicki,J., Ozersky,P., Bielicki,L., Scott,K., Holmes,A.,
            Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
            Dauphin-Kohlberg,S., Kozlowicz-Reilly,A., Leonard,S., Rohlfing,T.,
            Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
            Strowmatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
            Woessner,J.P., Wendl,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,
```

```
            Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohldmann,P.E.,
            Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A.,
            Mardis,E.R., Clifton,S.W., Chissoe,S.L., Marra,M.A., Raymond,C.,
            Haugen,E., Gillett,W., Zhou,Y., James,R., Phelps,K., Iadanoto,S.,
            Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J.,
            Furey,T.S., Baertsch,R.A., Brent,M.R., Keibler,E., Flicek,P.,
            Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrents,D.,
            Chinwalla,A.T., Gish,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V.,
            Eichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K.
  TITLE     The DNA sequence of human chromosome 7
  JOURNAL   Nature 424 (6945), 157-164 (2003)
  MEDLINE   22737999
   PUBMED   12853948
REFERENCE   2  (bases 1 to 136901)
  AUTHORS   Bauer,H., Haakenson,B. and Nguyen,C.
  TITLE     The sequence of Homo sapiens BAC clone RP11-95P9
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 136901)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 136901)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 136901)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 136901)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 136901)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2004) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 16, 2001 this sequence version replaced gi:11597125.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0095P09
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is CTD-2023N18, 200 bp overlap the
            clone sequenced to the right is CTA-250D13, 200 bp overlap.  Actual
            start of this clone is at base position 127066 of CTD-2023N18
            actual end is at base position 26937 of CTA-250D13.
FEATURES             Location/Qualifiers
     source          1..136901
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-95P9"
                     /clone_lib="RPCI-11"
     repeat_region   545..1149
                     /rpt_family="L1"
     repeat_region   1195..1380
                     /rpt_family="L1"
     repeat_region   1533..1790
                     /rpt_family="L1"
     repeat_region   2972..3112
                     /rpt_family="MER1_type"
     repeat_region   3118..3324
                     /rpt_family="MER1_type"
     repeat_region   6682..6885
                     /rpt_family="MIR"
     repeat_region   7038..7388
                     /rpt_family="MaLR"
     repeat_region   7433..7770
                     /rpt_family="MaLR"
     repeat_region   7786..7823
                     /rpt_family="AT_rich"
     repeat_region   7829..7985
                     /rpt_family="(TA)n"
     repeat_region   8004..8064
                     /rpt_family="(TATG)n"
     repeat_region   8067..8206
                     /rpt_family="MaLR"
     repeat_region   9047..9246
                     /rpt_family="MaLR"
     repeat_region   9247..9288
                     /rpt_family="(TCTCTG)n"
     repeat_region   9289..9399
                     /rpt_family="MaLR"
     repeat_region   9412..9494
                     /rpt_family="MaLR"
     repeat_region   9495..9854
                     /rpt_family="MaLR"
     repeat_region   9855..9928
                     /rpt_family="MaLR"
     repeat_region   10530..10814
                     /rpt_family="Alu"
     repeat_region   10815..10909
                     /rpt_family="(CATA)n"
     repeat_region   14398..14434
                     /rpt_family="(TAA)n"
     repeat_region   14453..14659
                     /rpt_family="MIR"
     repeat_region   15191..15248
                     /rpt_family="AcHobo"
     repeat_region   15255..15456
                     /rpt_family="MaLR"
     repeat_region   15457..15494
                     /rpt_family="AT_rich"
     repeat_region   15515..15552
                     /rpt_family="AT_rich"
     repeat_region   15997..16043
                     /rpt_family="(CA)n"
     repeat_region   16497..16517
                     /rpt_family="AT_rich"
     repeat_region   16886..17023
                     /rpt_family="MER1_type"
     repeat_region   17189..17380
                     /rpt_family="MER1_type"
     repeat_region   17513..17684
                     /rpt_family="MIR"
     repeat_region   18387..18555
                     /rpt_family="MIR"
     repeat_region   18604..19120
                     /rpt_family="MaLR"
     repeat_region   20362..20667
                     /rpt_family="Alu"
     repeat_region   20943..20966
                     /rpt_family="AT_rich"
  repeat_region   21716..21879
                     /rpt_family="MIR"
     repeat_region   22121..22550
                     /rpt_family="ERVL"
     repeat_region   22576..23085
                     /rpt_family="MaLR"
     repeat_region   23097..24931

  Query Match             89.9%;  Score 571;  DB 9;  Length 136901;
  Best Local Similarity   93.7%;  Pred. No. 5.4e-169;
  Matches  595;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db      95677 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAACTGTAAAACTA 95618

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| ||| |||||||||||| ||||||| |||||||||||||||||||||||||||||
Db      95617 CAAATGGTTTTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACCGTGGACCC 95558

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||| |  ||||||||||||||| ||| |||||||||||  | ||||||||||||||||||
Db      95557 CTGAAATGGCCTGCTAGCCCATTCTCTGATGTTAATGATGTCGAAGGCACCCCTCCCGAG 95498

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db      95497 AAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGT 95438

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||| |||||||||||||||| ||||||||||||||||||||||||||| ||||||||||
Db      95437 CATCGGCCAACCTCCCCAACAACACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAGAGAC 95378

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db      95377 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 95318

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||||||||| ||||||||| ||| |||||||||||| ||||| ||||||||||||||||
Db      95317 GCATCCACCTTTAAACATGGTGCTCGCAACTTAGCTCCCACCCAACCAATCAGAGAGCTC 95258

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||| ||||||| |||||||||||||||||||||||||||||||||||
Db      95257 ACTAAAATGCTAATTATGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 95198

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||| |||||||||||||||||||| ||||||||||||||||||||| ||||| |
Db      95197 AGAGCACAGTGGGAGGGACAAGGATCGGGACATAAACCCAGGCATTCGAGCCAGCAACAG 95138
```

```
Qy          541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
                |||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||||
Db        95137 CAACCCCCTTTGGGTCCCCTCCATTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 95078

Qy          601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAAA 635
                ||||||||||||||||| |||||| ||||| ||||
Db        95077 ATTAAATCTTGCAACTGCAAAAAATAAAAATAAAA 95043


RESULT 11
AC068898/c
LOCUS       AC068898               176188 bp    DNA     linear   PRI 29-AUG-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-534L6, complete sequence.
ACCESSION   AC068898
VERSION     AC068898.8  GI:22539070
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176188)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176188)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 176188)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Aug 29, 2002 this sequence version replaced gi:14625082.
FEATURES             Location/Qualifiers
     source          1. .176188
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-534L6"
                     /clone_lib="RPCI-11"
ORIGIN

  Query Match             89.2%;  Score 566.4;  DB 9;  Length 176188;
  Best Local Similarity   93.5%;  Pred. No. 1.6e-167;
  Matches  591;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy            1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
                |||||||||||||||| ||||||||||||||| ||||||||||| |||||||||||||||
Db        57896 CCCTGTATCTTTAACCGCCTTGTTAAGTTTGTGTCTTCCAGAATTAAAACTGTAAAACTA 57837

Qy           61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                ||||| |||||||||||||| | ||||||| |||||||| ||||||||  ||| ||||||
Db        57836 CAAATCGTTCTTCAAATGGAACCCCAGATGCAGTCCATGTCTAAGATCTCCCGCGGACCC 57777

Qy          121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
                || ||| ||||||||||||||||||| | |||||||||||| |||| ||| ||||| |||
Db        57776 CTAGACTGGCCTGCTAGCCCATGCTCTGGTGTTAATGACATCGAAGTCACTCCTCCTGAG 57717

Qy          181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
                |||||||||||||||||||||||||||||||||| |||||| |||||||||||||| |||
Db        57716 GAAATCTCAACTGCACAACCCCTACTATGCCCCAGTTCAGCAGGAAGCAGTTAGAGTGGT 57657

Qy          241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
                |||||||||||||| |||||||||||||||||||||||||||||| |||||||||||||
Db        57656 CATCAGCCAACCTCACCAACAGCACTTGGGTTTTCCTGTTGAGAGAGGGACTGAGAGAC 57597

Qy          301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGAAGGTGACT 360
                |||||||||||||||||||||||| |  |||||||||||| ||||||||||||||||| 
Db        57596 AGGACTAGCTGGATTTCCTAGGCCGATTAAGAATCCCTAAACCTAGCTGGAAGGTGACC 57537
```

```
Qy          361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                |||||||||| |||||| ||||||||||||||||||||||||| ||||||||||||||||
Db        57536 GCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAACCAATCAGAGAGCTC 57477

Qy          421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
                |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        57476 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 57417

Qy          481 AGAGCACAGCGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
                ||||||||| |||||||||||||| |  ||||||||||||||||||||||| |||||||
Db        57416 AGAGCACAGTGGAGGGACAAGGATTGCAATATAAACCCAGGCATTCGAGCCAGCAACGG 57357

Qy          541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
                ||||| |||||||||||| |||||||||||||| ||||||||||||||||||||||||||
Db        57356 CAACCGCCTTTGGGTCCCTTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 57297

Qy          601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAA 632
                ||||||||||||||||||||||||||||||||
Db        57296 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAA 57265


RESULT 12
AL583805
LOCUS       AL583805                89728 bp    DNA     linear   PRI 29-SEP-2001
DEFINITION  Human DNA sequence from clone RP11-134K1 on chromosome 9, complete
            sequence.
ACCESSION   AL583805 AC051658
VERSION     AL583805.7  GI:15865009
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 89728)
  AUTHORS   Clark,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 2, 2001 this sequence version replaced gi:14702155.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-134K1 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-134K1 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-134K1 is at 89728 in this
            sequence. The true right end of clone RP11-60C15 is at 2000 in this
            sequence.
```

```
FEATURES             Location/Qualifiers
     source          1..89728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-134K1"
                     /clone_lib="RPCI-11.1"
ORIGIN

  Query Match             88.9%;  Score 564.6;  DB 9;  Length 89728;
  Best Local Similarity   94.2%;  Pred. No. 5.4e-167;
  Matches  598;  Conservative    0;  Mismatches   34;  Indels    3;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||| |||||||||||||||| || ||||||   ||
Db      83434 CCCTGTATCTTTAACCTCCTTGTTAAGTGTGTCTCTTCCAGAATCGAAGCTGTAA---TA 83490

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
               | || ||||||||||||||||  |||||| ||||||||||||||||| |||| ||||||
Db      83491 TATATGGTTCTTCAAATGGAGCCACAGATGCAGTCCATGACTAAGATCTACCGCGGACCC 83550

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||| |||||||||||||||||||| |||
Db      83551 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAACGACATTGAAGGCACCCCTCCTGAG 83610

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||  |||||||||||||||||||| ||||||||||||||||||
Db      83611 GAAATCTCAACTGCACAACAACTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 83670

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
                ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db      83671 TGTCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGCGGGACTGAGAGAC 83730

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  |  ||||||||||||||||||||||||||||||||
Db      83731 AGGACTAGCTGGATTTCCTAGGCGGAATAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 83790

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||||||||| |||||| ||||||||||||||||||||||||| ||||||||||||||||
Db      83791 GCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAACCAATCAGAGAGCTC 83850

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      83851 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 83910

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| ||||||||||||||| ||||||||||||||||||||||||||||||| |
Db      83911 AGAGCACAGCAGGAGGGACAAGGATCAGGATATAAACCCAGGCATTCGAGCCGGCAACAG 83970

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| |||||||||||| |||||||||||||
Db      83971 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATTCTATTTCACTCT 84030

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||| |||||||| |||||||||||||||||
Db      84031 ATTAAATCCTGCAACTGCAAAAAAAAAAAAAAAAA 84065


RESULT 13
AC072023
LOCUS       AC072023              112405 bp    DNA     linear   PRI 16-JAN-2003
DEFINITION  Homo sapiens 3 BAC RP11-3J2 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC072023
VERSION     AC072023.9  GI:27764634
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112405)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 112405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 112405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 112405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 112405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 112405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 16, 2003 this sequence version replaced gi:21539113.
```

```
          INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
          gc-help@bcm.tmc.edu

          CLONE LENGTH: This sequence does not necessarily represent the
          entire insert of this clone. Overlapping regions of clones are only
          sequenced and submitted once, so the sequence for the remainder of
          the insert may be found in the record for the adjacent clones.
          Overlapping clones are noted at the beginning and end of the
          Features listing.

          ANNOTATION OF FEATURES:
             STSs are identified using ePCR (Genome Res. 7:541-550) searches
          of a local database that includes entries from dbSTS, GDB, and
          local mapping efforts.
             Repeats are identified using RepeatMasker (A. Smit and P. Green,
          unpublished.) for Human and Mouse sequences.
             Genes and Region of sequence similarity are identified by BLAST
          (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
          EST and cDNA sequences. Genes demonstrate at least two exons
          flanked by consensus splice sites that maintained sequence
          continuity across the splice junctions. Sequences that are not
          identical matches are annotated as similar.

          SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
          standard of double strand coverage with a minimum of 2 clones and 2
          reads with no ambiguities or 2 chemistries with a minimum of 2
          clones and 3 reads with no ambiguities. If the sequence quality for
          a region does not meet this standard, it will be indicated in the
          annotation as Low Coverage.

          QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
          standards - estimated error rate less than 1 per 10,000 bases.
          Reports of lowest quality individual bases and measures of base
          quality are listed below. Description of the metrics can be found
          at URL:
          http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
          ml.
FEATURES             Location/Qualifiers
     source          1..112405
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-3J2"
     misc_feature    1..2000
                     /note="overlaps bases 38072..40071 of clone AC125608"
                     /function="clone overlap"
     repeat_region   complement(1..261)
                     /rpt_family="L1M4"
     repeat_region   complement(431..913)
                     /rpt_family="L1ME"
     repeat_region   complement(914..1221)
                     /rpt_family="AluSg"
     repeat_region   complement(1222..1808)
                     /rpt_family="L1ME"
     repeat_region   complement(2045..2198)
                     /rpt_family="L1ME"
     repeat_region   2454..2811
                     /rpt_family="THE1C"
     repeat_region   2813..3092
                     /rpt_family="MLT1I"
     repeat_region   3194..3870
                     /rpt_family="L1PA3"
     repeat_region   3871..3891
                     /rpt_family="AT_rich"
     repeat_region   complement(4297..4671)
                     /rpt_family="MSTD"
     repeat_region   4714..4913
                     /rpt_family="MLT1I"
     repeat_region   5632..5670
                     /rpt_family="AT_rich"
     repeat_region   complement(6337..6630)
                     /rpt_family="AluSx"
     repeat_region   7063..7093
                     /rpt_family="(TTTA)n"
     repeat_region   complement(7094..7374)
                     /rpt_family="AluSp"
     repeat_region   complement(7386..7695)
                     /rpt_family="AluSx"
     repeat_region   7910..8035
                     /rpt_family="MIR"
     repeat_region   8824..10387
                     /rpt_family="L1PA16"
     repeat_region   10388..10412
                     /rpt_family="AT_rich"
     repeat_region   10637..10675
                     /rpt_family="AT_rich"
     repeat_region   10754..10818
                     /rpt_family="L1PB1"
     repeat_region   complement(10830..10905)
                     /rpt_family="L1PA16"
     repeat_region   complement(10956..11184)
                     /rpt_family="L1"
     repeat_region   11295..11658
                     /rpt_family="L2"
     repeat_region   complement(12489..12812)
                     /rpt_family="L1PA15"
     repeat_region   complement(12819..13533)
                     /rpt_family="L1"
     repeat_region   13599..13619
                     /rpt_family="AT_rich"
     repeat_region   14620..14933
                     /rpt_family="AluSg"
     repeat_region   complement(15183..15254)
                     /rpt_family="L2"
     repeat_region   complement(16463..17175)
                     /rpt_family="L1P"
     repeat_region   17173..18635
                     /rpt_family="L1PA7"
     repeat_region   18793..19462
                     /rpt_family="L1M4"
     repeat_region   19461..20130
                     /rpt_family="L1ME"
     repeat_region   20776..21073
                     /rpt_family="AluSx"

  Query Match             88.2%;  Score 560.2;  DB 9;  Length 112405;
  Best Local Similarity   93.8%;  Pred. No. 1.4e-165;
  Matches  594;  Conservative    0;  Mismatches   38;  Indels    1;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db      67581 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAATGGTAAAACTA 67640

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| | |||||||||||||| ||||||| |||||||||||||||||||||   |||||
Db      67641 CAAATGGGTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACCACAGACCC 67700

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||| |||||||||||||||| |||||||||||||||||  |||||||||||||| ||
Db      67701 CTGGACTGGCCTGCTAGCCCATGATCCGATGTTAATGACATCAAAGGCACCCCTCCCAAG 67760

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||| |||||||||||||||||||||||||| ||||||||||||||| ||
Db      67761 GAAATCTCAACTGCCCAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 67820

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||  |||||||||||||||||||||||| | ||||||||||||||||||||||||||| 
Db      67821 CATTGGCCAACCTCCCCAACAGCACTTGGATCTTCCTGTTGAGAGGGGGGACTGAGAGAT 67880

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||| 
Db      67881 AGGACTAGCTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 67940

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||||||||| |||||| |||||| |||||||||||||||||| ||||||||||||||||
Db      67941 GCATCCACCTTTAAACACGGGGCTCGCAACTTAGCTCACACCCAACCAATCAGAGAGCTC 68000

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| ||||||||||  |||||||||||||||||||||||
Db      68001 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAA-TAATAGCCAATCATCTATTGCCTG 68059

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||||||||||||||||| ||||||||||| |||||||||||||| ||||| ||
Db      68060 AGAGCACAGCGGGAGGGACAAGGAACGGGATATAAAACCAGGCATTCGAGCTGGCAATGG 68119

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              |||||||||||||||||||| || ||||||||| ||||||||||||||||||||||||||
Db      68120 CAACCCCCTTTGGGTCCCCTTCCCTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 68179

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAA 633
              ||||||||||||||||| |||||||||||||||
Db      68180 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAA 68212


RESULT 14
BS000045/c
LOCUS       BS000045               179114 bp    DNA     linear   PRI 12-JUN-2004
DEFINITION  Pan troglodytes chromosome 22 clone:RP43-179P23, map 22, complete
            sequences.
ACCESSION   BS000045 BA000046
VERSION     BS000045.1  GI:37537312
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   The International Chimpanzee Chromosome 22 Consortium.
  TITLE     DNA sequence and comparative analysis of chimpanzee chromosome 22
  JOURNAL   Nature 429, 382-388 (2004)
REFERENCE   2  (bases 1 to 179114)
  AUTHORS   Hattori,M., Toyoda,A., Watanabe,H., Taylor,T.D., Kuroki,Y.,
            Fujiyama,A. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
            *Chinese National Human Genome Center at Shanghai, Shanghai, China;
            *GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
            of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
            Center, Daejeon, Korea;
            *Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
            *National Institute of Genetics, Mishima, Japan;
            *National Yang Ming University Genome Research Center, Taipei,
            Taiwan;
            *RIKEN Genomic Sciences Center, Yokohama, Japan.
            -------------- Genome Center
            Center: RIKEN Genomic Sciences Center
            Center code: RIKEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
            Center project name:The Chimpanzee Chromosome 22 Sequencing Project
            Center clone name: RP43-179P23
            -------------- Summary Statistics
            Sequencing vector: pUC18,pUC13,pTZ19R; 100% of reads Chemistry:
            Dye-terminator Big Dye and ET; 100% of reads Assembly program:
            Phrap; version 0.990329
            Consensus quality:   178,810 bases at least Q40
            Consensus quality:       263 bases at least Q30
            Consensus quality:        35 bases at least Q20
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30);
            an attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by at one
            plasmid
            subclone or more than one M13 subclone;
            and the assembly was confirmed by restriction digest.
            --------------
            Source information:
            The RPCI-43 chimpanzee BAC library was prepared from DNA isolated
            from the blood of a single male chimpanzee using published
            protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the
            chimpanzee ('Clint') was obtained from the Yerkes Primate Center in
            Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
            Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
            characteristics are described at
            http://www.chori.org/bacpac/mchimp43.htm.
            The clone may be obtained from Pieter J. de Jong and coworkers
            (http://www.chori.org/bacpac).
            VECTOR:  pBACe3.6
            The CHORI-251 chimpanzee BAC library was prepared from DNA isolated
            from the blood of a single male chimpanzee using published
            protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the
            chimpanzee ('Clint') was obtained from the Yerkes Primate Center in
            Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
            Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
            characteristics are described at
            http://www.chori.org/bacpac/chimpanzee251.htm.
            The clone may be obtained from Pieter J. de Jong and coworkers
            (http://www.chori.org/bacpac).
            VECTOR:  pTARBAC2.1
            The PTB1 chimpanzee BAC library was prepared from DNA isolated from
            cultured cells established from the blood of a single male
            chimpanzee.
            Clones may be obtained from Asao Fujiyama and co-workers
            (http://www.gsc.riken.go.jp).
            VECTOR:  pKS145
            The PTF22 chimpanzee Fosmid library was prepared from DNA isolated
            from cultured cells established from the blood of a single male
            chimpanzee.
            Clones may be obtained from Asao Fujiyama and co-workers
            (http://www.gsc.riken.go.jp).
            VECTOR:  pKS143
            --------------
            Sequence Quality Assessment:
            This entry has been annotated with sequence
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than 1 error in
            10,000 bp.
            ----------------------
            Neighboring clones: RP43-014B20(left) and PTB-042H12(right).
FEATURES             Location/Qualifiers
     source          1..179114
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="22"
                     /clone="RP43-179P23"
                     /clone_lib="RPCI-43 chimpanzee BAC"
ORIGIN

  Query Match             88.2%;  Score 559.8;  DB 9;  Length 179114;
  Best Local Similarity   94.0%;  Pred. No. 2e-165;
  Matches  582;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     169185 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 169126

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| |||||||||||||||||||||| || |||
Db     169125 CAAATAGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACCGCGGGCCC 169066
```

```
Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              || ||| ||||||||||||||||||||||| |||||||||| |||| |||||||||||||
Db     169065 CTAGACTGGCCTGCTAGCCCATGCTCCGATATTAATGACATCGAAGACACCCCTCCCGAG 169006

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||| |||||||||||||||||||||||||| ||||||||||||||| ||
Db     169005 GAAATCTCAACTGCCCAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGCCGT 168946

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||  |||||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db     168945 CATTGGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGGGGGGGTACTGAGAGAC 168886

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||| ||||| |||||||||||| ||||||| ||||
Db     168885 AGGACTAGCTGGATTTCCTAGGCCAACTAAGAAGCCCTAAGCCTAGATGGGAAGATGACC 168826

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||||||| | ||||||||||||| |||||||||||||||||| |||| |||||||||||
Db     168825 GCATCCACTTTTAAACATGGGGCTCGCAACTTAGCTCACACCCAACCAGTCAGAGAGCTC 168766

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     168765 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 168706

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| ||||||||||| ||| |||||||||||||||||||| ||| ||||||||
Db     168705 AGAGCACAGCAGGAGGGACAAGCATCAGGATATAAACCCAGGCATTCAAGCTGGCAACGG 168646

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||| |||||
Db     168645 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGGGCTCTGTTTTCACTCTATTTAACTCT 168586

Qy        601 ATTAAATCTTGCAACTGAA 619
              | ||||||||||||||| |
Db     168585 ACTAAATCTTGCAACTGCA 168567
```

```
RESULT 15
AP001538/c
LOCUS       AP001538               174019 bp    DNA     linear   PRI 25-MAR-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B853K11,
            LL56-APP region, complete sequence.
ACCESSION   AP001538
VERSION     AP001538.1  GI:7328982
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174019)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 174,019 genomic DNA of 21q21.1-q21.2
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 174019)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
FEATURES             Location/Qualifiers
     source          1..174019
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.1-q21.2"
                     /clone="B853K11"
ORIGIN
```

```
  Query Match             87.7%;  Score 556.6;  DB 9;  Length 174019;
  Best Local Similarity   93.7%;  Pred. No. 2.1e-164;
  Matches  580;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36710 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 36651

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| ||||||||||||| || ||||||| |||||||||||||||||||||| || |||
Db      36650 CAAATAGTTCTTCAAATGGGGCCCCAGATGCAGTCCATGACTAAGATCCACCGCGGGCCC 36591

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              || ||| ||||||||||||||||||||||| |||||||||| ||||||||||||||| ||
Db      36590 CTAGACTGGCCTGCTAGCCCATGCTCCGATATTAATGACATCGAAGGCACCCCTCCCAAG 36531

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||| ||||||| |||||||||||||||||||||||||| ||||||||||||||| ||
Db      36530 GAAATCACAACTGCCCAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGCTGT 36471

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||  |||||||||||||||||||||||||||||||||||||| | ||| ||||||||||
Db      36470 CATTGGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGGGCGGGTACTGAGAGAC 36411

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db      36410 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 36351

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||||||| | ||||||||||||| |||||||||||||||||| |||| |||||||||||
Db      36350 GCATCCACTTTTAAACATGGGGCTCGCAACTTAGCTCACACCCAACCAGTCAGAGAGCTC 36291

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      36290 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 36231

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| ||||||||||||||| |||| ||||||||||||||| ||| ||||||||
Db      36230 AGAGCACAGCAGGAGGGACAAGGATCAGGATGTAAACCCAGGCATTCAAGCTGGCAACGG 36171

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              |||||||||||||||||||| |||||||||||| ||||||||||||||||||||||||||
Db      36170 CAACCCCCTTTGGGTCCCCTGCCTTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 36111

Qy        601 ATTAAATCTTGCAACTGAA 619
              | ||||||||||||||| |
Db      36110 ACTAAATCTTGCAACTGCA 36092
```

```
RESULT 16
AP001674/c
LOCUS       AP001674               340000 bp    DNA     linear   PRI 21-MAY-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, section 18/105.
ACCESSION   AP001674 AL163219 BA000005
VERSION     AP001674.1  GI:7768666
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,K., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     The DNA sequence of human chromosome 21
  JOURNAL   Nature 405 (6784), 311-319 (2000)
  MEDLINE   20289799
   PUBMED   10830953
REFERENCE   2  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
COMMENT     On May 30, 2000 this sequence version replaced gi:7717271.
            The chromosome 21 mapping and sequencing consortium consisting of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: hattori@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Molecular Biology, * Tokyo
            160-8582, Japan,
            * e.mail: nshimizu@dmb-med.keio.ac.jp
            * URL: http://www.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/
            AL163219: Submitted (10-Apr-2000).
FEATURES             Location/Qualifiers
     source          1..340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
     source          <1..6149
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.1"
                     /clone="R749F18, 5' partial"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AP001537"
     source          <1..165231
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
                     /chromosome="21"
                     /map="21q21.1"
                     /clone="R853K11, 5' partial"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AP001538"
     source          136796..315056
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.1"
                     /clone="R845I12"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AF245226"
     source          156433..277809
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.1"
                     /clone="P646P19"
                     /clone_lib="RPCI1,3-5 PAC library"
                     /note="Accession No. AL157359"
     source          207594..250588
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.1"
                     /clone="Q3D19"
                     /clone_lib="LL21NC02-Q Cosmid library"
                     /note="Accession No. AL049753"
     source          276958..301481
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.1"
                     /clone="R755I19"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AP001345"
     source          299634..>340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.1"
                     /clone="R28M9, 3' partial"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AP000431"
     source          311751..>340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.1"
                     /clone="c103E2331, 3' partial"
                     /clone_lib="ICRFc103 Cosmid library"
                     /note="Accession No. AL132993"
     repeat_region   15..444
                     /note="L1MA4A"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   555..3531
                     /note="L1MA4A"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   3770..3876
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(5645..6178)
                     /note="L1MA2"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   6181. .6534
                     /note="THE1B"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   8021. .8370
                     /note="THE1B"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   complement(8373. .10032)
                     /note="L1MA2"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   10213. .10520
                     /note="AluJb"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   15810. .15851
                     /note="(CA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   complement(16628. .16697)
                     /note="MER53"
                     /rpt_family="DNA"
                     /rpt_type=DISPERSED
     repeat_region   16638. .16715
                     /note="MER53"
                     /rpt_family="DNA"
                     /rpt_type=DISPERSED
     repeat_region   18847. .19585
                     /note="L2"
                     /rpt_family="LINE/L2"

  Query Match             87.7%;  Score 556.6;  DB 9;  Length 340000;
  Best Local Similarity   93.7%;  Pred. No. 2.4e-164;
  Matches  580;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      27921 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 27862

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| ||||||||||||| || ||||||| |||||||||||||||||||||| || |||
Db      27861 CAAATAGTTCTTCAAATGGGGCCCCAGATGCAGTCCATGACTAAGATCCACCGCGGGCCC 27802

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              || ||| ||||||||||||||||||||||| |||||||||| ||||||||||||||| ||
Db      27801 CTAGACTGGCCTGCTAGCCCATGCTCCGATATTAATGACATCGAAGGCACCCCTCCCAAG 27742

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||| ||||||| |||||||||||||||||||||||||| ||||||||||||||| ||
Db      27741 GAAATCACAACTGCCCAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGCTGT 27682

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||  |||||||||||||||||||||||||||||||||||||| | ||| ||||||||||
Db      27681 CATTGGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGGGCGGGTACTGAGAGAC 27622

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||| 
Db      27621 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 27562

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||||||| | ||||||||||||| |||||||||||||||||| |||| |||||||||||
Db      27561 GCATCCACTTTTAAACATGGGGCTCGCAACTTAGCTCACACCCAACCAGTCAGAGAGCTC 27502

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      27501 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 27442

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| ||||||||||||||| |||| ||||||||||||||| ||| ||||||||
Db      27441 AGAGCACAGCAGGAGGGACAAGGATCAGGATGTAAACCCAGGCATTCAAGCTGGCAACGG 27382

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              |||||||||||||||||||| |||||||||||| ||||||||||||||||||||||||||
Db      27381 CAACCCCCTTTGGGTCCCCTGCCTTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 27322

Qy        601 ATTAAATCTTGCAACTGAA 619
              | ||||||||||||||| |
Db      27321 ACTAAATCTTGCAACTGCA 27303


RESULT 17
AP001545
LOCUS       AP001545              149755 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-762G24 map.18q21, WORKING
            DRAFT SEQUENCE, 13 unordered pieces.
ACCESSION   AP001545
VERSION     AP001545.3  GI:9188503
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149755)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 149,755 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 149755)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On Jul 14, 2000 this sequence version replaced gi:8117383.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-762G24
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 145596 bases at least Q40
                Consensus quality: 147102 bases at least Q30
                Consensus quality: 147971 bases at least Q20
                Insert size: 148555; sum-of-contigs
                Quality coverage: 10.98x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            13 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                     1     44828 contig of  44828 bp in length
                 44929     82602 contig of  37674 bp in length
                 82703    101344 contig of  18642 bp in length
                101445    116205 contig of  14761 bp in length
                116306    123784 contig of   7479 bp in length
                123885    130355 contig of   6471 bp in length
                130456    133619 contig of   3164 bp in length
                133720    137416 contig of   3697 bp in length
                137517    141540 contig of   4024 bp in length
                141641    144351 contig of   2711 bp in length
                144452    146667 contig of   2216 bp in length
            146768     148219 contig of    1452 bp in length
            148320     149755 contig of    1436 bp in length.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 13 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *          1     44828: contig of 44828 bp in length
          *      44829     44928: gap of 100 bp
          *      44929     82602: contig of 37674 bp in length
          *      82603     82702: gap of 100 bp
          *      82703    101344: contig of 18642 bp in length
          *     101345    101444: gap of 100 bp
          *     101445    116205: contig of 14761 bp in length
          *     116206    116305: gap of 100 bp
          *     116306    123784: contig of 7479 bp in length
          *     123785    123884: gap of 100 bp
          *     123885    130355: contig of 6471 bp in length
          *     130356    130455: gap of 100 bp
          *     130456    133619: contig of 3164 bp in length
          *     133620    133719: gap of 100 bp
          *     133720    137416: contig of 3697 bp in length
          *     137417    137516: gap of 100 bp
          *     137517    141540: contig of 4024 bp in length
          *     141541    141640: gap of 100 bp
          *     141641    144351: contig of 2711 bp in length
          *     144352    144451: gap of 100 bp
          *     144452    146667: contig of 2216 bp in length
          *     146668    146767: gap of 100 bp
          *     146768    148219: contig of 1452 bp in length
          *     148220    148319: gap of 100 bp
          *     148320    149755: contig of 1436 bp in length.
FEATURES             Location/Qualifiers
     source          1..149755
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q21"
                     /clone="RP11-762G24"
     misc_feature    1..44828
                     /note="assembly_fragment"
     misc_feature    44929..82602
                     /note="assembly_fragment"
     misc_feature    82703..101344
                     /note="assembly_fragment"
     misc_feature    101445..116205
                     /note="assembly_fragment"
     misc_feature    116306..123784
                     /note="assembly_fragment"
     misc_feature    123885..130355
                     /note="assembly_fragment"
     misc_feature    130456..133619
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    133720..137416
                     /note="assembly_fragment"
     misc_feature    137517..141540
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    141641..144351
                     /note="assembly_fragment"
     misc_feature    144452..146667
                     /note="assembly_fragment"
     misc_feature    146768..148219
                     /note="assembly_fragment"
     misc_feature    148320..149755
                     /note="assembly_fragment"
ORIGIN

  Query Match             86.9%;  Score 551.8;  DB 2;  Length 149755;
  Best Local Similarity   91.8%;  Pred. No. 6.8e-163;
  Matches  583;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      67067 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACTA 67126

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||  ||||||||||||||| ||||||| ||||||||||||||||| |||   |||||
Db      67127 CAAATCATTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCACAGACCC 67186

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||| ||||||||||||||||||||||||||||||||| ||||||| ||||||| ||
Db      67187 CTGGACCAGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCAACCCTCCCAAG 67246

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||||||||| |||||| |||||||||||| |||||
Db      67247 GAAATCTCAACTGCACAACCCCTACTATGCCCCAGTTCAGCAGGAAGCAGTTAGGGCGGT 67306

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db      67307 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTCAGAGGGGGTACTGAGAGAC 67366

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||| |||||||||||||||||||
Db      67367 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAACCTAGCTGGGAAGGTGACT 67426

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||| ||||| |||||| | |||||||||||||||||||||| |||||||||||||||||
Db      67427 GCATTCACCTTTAAACACGAGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 67486

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      67487 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 67546

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||| |||||||||||||||| |||||||||||||||||||||||||||||||||
Db      67547 AGAGCACAGTGGGAGGGACAAGGATCAGGATATAAACCCAGGCATTCGAGCCGGCAACGG 67606

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||  |   |
Db      67607 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTAAATCTT 67666

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
                 |   |    ||   ||| |||||||  |||||
Db      67667 TCAACTGCAAAAAATAAAAATAAAAAAATTAAAAA 67701


RESULT 18
AC090313/c
LOCUS       AC090313              152980 bp    DNA     linear   HTG 27-MAR-2003
DEFINITION  Homo sapiens chromosome 18 clone RP11-762G24 map 18.
ACCESSION   AC090313
VERSION     AC090313.12  GI:22123575
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152980)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-762G24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152980)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 152980)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 6, 2002 this sequence version replaced gi:21699527.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12582
                Center clone name: 762_G_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1   152980: contig of 152980 bp in length.
FEATURES             Location/Qualifiers
     source          1..152980
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-762G24"
                     /clone_lib="RPCI-11 Human Male BAC"
ORIGIN

  Query Match             86.9%;  Score 551.8;  DB 2;  Length 152980;
  Best Local Similarity   91.8%;  Pred. No. 6.8e-163;
  Matches  583;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;
```

```
Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      48688 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACTA 48629

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||  ||||||||||||||| ||||||| ||||||||||||||||| |||   |||||
Db      48628 CAAATCATTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCACAGACCC 48569

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||| ||||||||||||||||||||||||||||||||| ||||||| ||||||| ||
Db      48568 CTGGACCAGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCAACCCTCCCAAG 48509

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||||||||| |||||| |||||||||||| |||||
Db      48508 GAAATCTCAACTGCACAACCCCTACTATGCCCCAGTTCAGCAGGAAGCAGTTAGGGCGGT 48449

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db      48448 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTCAGAGGGGGTACTGAGAGAC 48389

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||| |||||||||||||||||||
Db      48388 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAACCTAGCTGGAAGGTGACT 48329

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||| ||||| |||||| | |||||||||||||||||||||| |||||||||||||||||
Db      48328 GCATTCACCTTTAAACACGAGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 48269

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      48268 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 48209

Qy        481 AGAGCACAGCGGAGGGACAAGGATCGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||| ||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      48208 AGAGCACAGTGGAGGGACAAGGATCAGGATATAAACCCAGGCATTCGAGCCGGCAACGG 48149

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||  |   |
Db      48148 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTAAATCTT 48089

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
                 |   |    ||   ||| |||||||  |||||
Db      48088 TCAACTGCAAAAAATAAAAATAAAAAAATTAAAAA 48054
```

```
RESULT 19
AC093531/c
LOCUS       AC093531          163803 bp    DNA     linear   PRI 16-NOV-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
ACCESSION   AC093531
VERSION     AC093531.2  GI:16945981
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163803)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163803)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 163803)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15383820.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.3.
FEATURES             Location/Qualifiers
     source          1..163803
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-405L7"
ORIGIN

  Query Match             86.9%;  Score 551.8;  DB 9;  Length 163803;
  Best Local Similarity   93.2%;  Pred. No. 6.9e-163;
  Matches  577;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||| ||||||||||||||||||||||||||||  || ||||||| |||
Db     117348 CCCTGTATCTTTAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 117289

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db     117288 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 117229

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||| | |||||| ||||||||||||||||||| |||||||||||||
Db     117228 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 117169

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db     117168 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 117109

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db     117108 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 117049

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db     117048 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 116989

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db     116988 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 116929

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||| ||||||||| ||||||||||| |||||||||||||||||||||||
Db     116928 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 116869

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||||||| |||||||||| ||||||||||||||||||||| |||| ||||| |
Db     116868 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACCCAGGCATTCAAGCCAGCAACAG 116809

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db     116808 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 116749

Qy        601 ATTAAATCTTGCAACTGAA 619
              ||||||||||||||||| |
Db     116748 ATTAAATCTTGCAACTGCA 116730


RESULT 20
AC107075/c
LOCUS       AC107075               176425 bp    DNA     linear   PRI 12-JUN-2002
DEFINITION  Homo sapiens BAC clone RP11-452N17 from 2, complete sequence.
ACCESSION   AC107075
VERSION     AC107075.4  GI:21263352
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176425)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 176425)
  AUTHORS   Swearengen-Shahid,S., Meyer,R. and Dignan,G.
  TITLE     The sequence of Homo sapiens BAC clone RP11-452N17
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 176425)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 176425)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 176425)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 30, 2002 this sequence version replaced gi:18640713.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0452N17
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-624F4; the clone sequenced
            to the right is RP11-362J3.  Actual start of this clone is at base
            position 1 of RP11-452N17; actual end is at base position 176425 of
            RP11-452N17.
            Polymorphisms exist between AC110086 and AC107075. Data from
            AC110086, AC116626 and AC018872 was used to finish AC107075.
FEATURES             Location/Qualifiers
     source          1..176425
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-452N17"
                     /clone_lib="RPCI-11"
     repeat_region   2423..2523
                     /rpt_family="MIR"
     repeat_region   3564..3598
                     /rpt_family="AT_rich"
     repeat_region   4522..4567
                     /rpt_family="AT_rich"
     repeat_region   4679..4756
                     /rpt_family="(TATAA)n"
     repeat_region   5700..5733
                     /rpt_family="(A)n"
     repeat_region   5973..6006
                     /rpt_family="(A)n"
     repeat_region   6123..6144
                     /rpt_family="AT_rich"
     repeat_region   6803..6896
                     /rpt_family="L1"
     repeat_region   6923..6999
                     /rpt_family="L1"
     repeat_region   7230..7715
                     /rpt_family="MaLR"
     repeat_region   7852..7966
                     /rpt_family="Tc2"
     repeat_region   7966..8026
                     /rpt_family="Tc2"
     repeat_region   8043..8172
                     /rpt_family="Tc2"
     repeat_region   8174..8247
                     /rpt_family="(TATATG)n"
     repeat_region   8250..8456
                     /rpt_family="Tc2"
     repeat_region   8922..9642
                     /rpt_family="L1"
     repeat_region   11590..11632
                     /rpt_family="AT_rich"
     repeat_region   12173..12321
                     /rpt_family="MER1_type"
     repeat_region   12323..12343
                     /rpt_family="(A)n"
     repeat_region   13107..14677
                     /rpt_family="L1"
     repeat_region   14678..14704
                     /rpt_family="(TTTG)n"
     repeat_region   14708..14996
                     /rpt_family="Alu"
     repeat_region   14997..15173
                     /rpt_family="L1"
     repeat_region   15615..15720
                     /rpt_family="CT-rich"
     repeat_region   15811..16031
                     /rpt_family="MaLR"
     repeat_region   16093..16330
                     /rpt_family="L1"
     repeat_region   16724..17718
                     /rpt_family="L1"
     repeat_region   18878..19038
                     /rpt_family="Alu"
     repeat_region   19298..19551
                     /rpt_family="Alu"
     repeat_region   19552..19912
                     /rpt_family="MaLR"
     repeat_region   23674..23705
                     /rpt_family="(CAGA)n"
     repeat_region   25248..25839
                     /rpt_family="MaLR"
     repeat_region   27499..27532
                     /rpt_family="AT_rich"
     repeat_region   27693..27840
                     /rpt_family="MER1_type"
     repeat_region   27902..28090
                     /rpt_family="MIR"
     repeat_region   28091..28254
                     /rpt_family="MER1_type"
     repeat_region   28283..28477
                     /rpt_family="MER1_type"
     repeat_region   28510..28533
                     /rpt_family="AT_rich"
     repeat_region   28629..28649
                     /rpt_family="AT_rich"
     repeat_region   29707..29751
                     /rpt_family="(TA)n"
     repeat_region   30414..30847
                     /rpt_family="L1"
     repeat_region   31385..31942
                     /rpt_family="L1"
     repeat_region   32421..32591
                     /rpt_family="L2"
     repeat_region   32592..32983
                     /rpt_family="MaLR"
     repeat_region   32984..33714
                     /rpt_family="L2"
     repeat_region   34919..35339
                     /rpt_family="MaLR"
     repeat_region   35752..36049
                     /rpt_family="Alu"
     repeat_region   36274..36330
                     /rpt_family="Alu"
     repeat_region   36767..36809
                     /rpt_family="AT_rich"
     repeat_region   37396..37437
                     /rpt_family="AT_rich"
     repeat_region   37816..37850
                     /rpt_family="A-rich"
     repeat_region   38954..39375
                     /rpt_family="L1"
     repeat_region   40130..40216
                     /rpt_family="L2"
     repeat_region   41723..42015
                     /rpt_family="Alu"

  Query Match             86.9%;  Score 551.8;  DB 9;  Length 176425;
  Best Local Similarity   92.8%;  Pred. No. 7e-163;
  Matches  593;  Conservative    0;  Mismatches   37;  Indels    9;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||| ||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db     161068 CCCTATATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACTA 161009

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||||||||||| |||||||||||
Db     161008 CAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCGTGGACCC 160949

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| ||
Db     160948 CTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATTGAAGGCACCCCTCCCAAG 160889

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||| ||||||||||||||||| |||| | ||||||||||||||||||
Db     160888 GAAATCTCAACTGCACCACCCCTACTATGCCCCAGTTCAACAGGAAGCAGTTAGAGCGGT 160829

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              | || |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     160828 CTTCGGCCAACCTCCCCAAAAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 160769

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||| || |||||||||||||||||||||||||||||||
Db     160768 AGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 160709

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCA-------- 412
              || |||| || |||||| ||||||| ||||||| |||||||  |||||||||
Db     160708 GCTTCCATCTTTAAACACGGGGCTTACAACTTAACTCACACATGACCAATCAGATAGTAA 160649

Qy        413 -GAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
               ||||||||||||||||||||||||||||| || |||||||||||||||||||||||| |
Db     160648 GGAGAGCTCACTAAAATGCTAATTAGGCAACAACAGGAGGTAAAGAAATAGCCAATCACC 160589

Qy        472 TATTGCCTGAGAGCACAGCGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              | ||||||||||||||||||||||||||| ||||||||||||||||||||||||| |||
Db     160588 TGTTGCCTGAGAGCACAGCGGAGGGACAATGATCGGGATATAAACCCAGGCATTCAAGC 160529

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              ||||||| |||||||||||||||||||||||||||||||||| |||||||||||||||||
Db     160528 CGGCAACAGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTA 160469

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAA 630
               |||||||||| |||||||||||||| ||||||||||||
Db     160468 CTTCACTCTATCAAATCTTGCAACTGCAAAAAAAAAAAA 160430

RESULT 21
AL139090/c
LOCUS       AL139090              128468 bp    DNA     linear   PRI 12-APR-2001
DEFINITION  Human DNA sequence from clone RP11-12B13 on chromosome 6 Contains
            STSs and GSSs, complete sequence.
ACCESSION   AL139090
VERSION     AL139090.11  GI:11228536
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128468)
  AUTHORS   Garner,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 20, 2000 this sequence version replaced gi:11125403.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-12B13 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-12B13 is at 1 in this sequence. The
            true left end of clone RP1-125N20 is at 128369 in this sequence.
            The true right end of clone RP11-13D18 is at 88802 in this
            sequence. This sequence was finished as follows unless otherwise
            noted: all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            RP11-12B13 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .128468
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-12B13"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .353
                     /note="match: GSS: Em:AQ609536"
     misc_feature    7. .448
                     /note="match: GSS: Em:AQ704228"
     misc_feature    68. .350
                     /note="match: GSS: Em:B75711"
     repeat_region   571. .1535
                     /note="LTR5 repeat: matches 1. .966 of consensus"
     repeat_region   1966. .2087
                     /note="L2 repeat: matches 2574. .2704 of consensus"
     repeat_region   2290. .2509
                     /note="L2 repeat: matches 2098. .2328 of consensus"
     repeat_region   2976. .3352
                     /note="L2 repeat: matches 1693. .2089 of consensus"
     repeat_region   3497. .3668
                     /note="MER20 repeat: matches 16. .188 of consensus"
     misc_feature    complement(4521. .5085)
                     /note="match: GSS: Em:AQ421033"
     misc_feature    complement(4571. .5086)
                     /note="match: GSS: Em:AQ883615"
     misc_feature    complement(4633. .5083)
                     /note="match: GSS: Em:AQ812769"
     repeat_region   5012. .5530
                     /note="MER41B repeat: matches 114. .601 of consensus"
     misc_feature    5081. .5634
                     /note="match: GSS: Em:AQ544837"
     repeat_region   6085. .6174
                     /note="MIR repeat: matches 64. .153 of consensus"
     repeat_region   6235. .6416
                     /note="MER5A repeat: matches 3. .186 of consensus"
     repeat_region   7146. .8805
                     /note="L1ME3 repeat: matches 4381. .6140 of consensus"
     repeat_region   8806. .9399
                     /note="L1PA5 repeat: matches 5526. .6143 of consensus"
     repeat_region   9400. .9710
                     /note="L1ME3 repeat: matches 4094. .4381 of consensus"
     repeat_region   9711. .15846
                     /note="L1PA6 repeat: matches 11. .6143 of consensus"
     repeat_region   15826. .16269
                     /note="L1M4 repeat: matches 3692. .4135 of consensus"
     repeat_region   17077. .17544
                     /note="L1M4 repeat: matches 2710. .3168 of consensus"
     repeat_region   17552. .17740
                     /note="L1MEc repeat: matches 2408. .2260 of consensus"
     repeat_region   17741. .18039
                     /note="AluSx repeat: matches 1. .299 of consensus"
     repeat_region   18040. .18346
                     /note="L1MEc repeat: matches 2106. .2409 of consensus"
     repeat_region   18447. .18874
                     /note="L1MEc repeat: matches 1469. .1910 of consensus"
     misc_feature    complement(18475. .18995)
                     /note="match: GSS: Em:AQ755181"
     repeat_region   18938. .19072
                     /note="L1MEc repeat: matches 1243. .1377 of consensus"
     repeat_region   19459. .19514
                     /note="28 copies 2 mer ca 96% conserved"
     repeat_region   20321. .20675
                     /note="LTR16B repeat: matches 97. .464 of consensus"
     repeat_region   20952. .21137
                     /note="AluJo repeat: matches 120. .307 of consensus"
     repeat_region   21203. .21254
                     /note="26 copies 2 mer aa 71% conserved"
     misc_feature    23289. .23682
                     /note="match: GSS: Em:AQ881909"
```

```
repeat_region   23421. .23472
                /note="MIR repeat: matches 90. .141 of consensus"
misc_feature    24395. .24791
                /note="match: GSS: Em:AQ027280"
repeat_region   24638. .24848
                /note="HAL1 repeat: matches 428. .638 of consensus"
repeat_region   25008. .25217
                /note="105 copies 2 mer aa 55% conserved"
repeat_region   25328. .25391
                /note="32 copies 2 mer ta 70% conserved"
repeat_region   25482. .26261
                /note="L1PA3 repeat: matches 5. .776 of consensus"
repeat_region   26257. .31504
                /note="L1PA3 repeat: matches 900. .6146 of consensus"
repeat_region   31614. .32312
                /note="L1M4 repeat: matches 3918. .4608 of consensus"
repeat_region   32398. .32650
                /note="L1M4 repeat: matches 4626. .4909 of consensus"
repeat_region   32799. .32963
                /note="FRAM repeat: matches -2. .162 of consensus"
repeat_region   33151. .33174
                /note="12 copies 2 mer tt 95% conserved"
repeat_region   33176. .33500
                /note="LTR17 repeat: matches 1. .326 of consensus"
repeat_region   33501. .33791
                /note="HERV17 repeat: matches 8244. .8523 of consensus"
repeat_region   33811. .38137
                /note="HERV17 repeat: matches 1651. .5933 of consensus"
repeat_region   38135. .39957
                /note="HERV17 repeat: matches 1. .1836 of consensus"
repeat_region   39958. .40489
                /note="LTR17 repeat: matches 257. .780 of consensus"
repeat_region   40490. .41321
                /note="L1ME3A repeat: matches 5261. .6131 of consensus"
repeat_region   41354. .41655
                /note="AluY repeat: matches 1. .302 of consensus"
repeat_region   41971. .42279
                /note="AluSx repeat: matches 1. .303 of consensus"
repeat_region   43255. .43365
                /note="L1M4 repeat: matches 2169. .2266 of consensus"
repeat_region   43366. .43774
                /note="MSTA repeat: matches 1. .426 of consensus"
repeat_region   43775. .44133
                /note="L1M4 repeat: matches 2266. .2691 of consensus"
repeat_region   44140. .44339
                /note="L1MA7 repeat: matches 6093. .6289 of consensus"
repeat_region   46709. .47216
                /note="L1MB8 repeat: matches 5658. .6169 of consensus"
repeat_region   47217. .47431
                /note="L1M4 repeat: matches 4637. .4856 of consensus"
repeat_region   47487. .47839
                /note="MLT2E repeat: matches 8. .345 of consensus"
repeat_region   47848. .48031
                /note="L1M4 repeat: matches 4451. .4645 of consensus"
repeat_region   48017. .49417
                /note="L1MEc repeat: matches 2373. .3522 of consensus"
repeat_region   49436. .50712
                /note="L1MD2 repeat: matches 5056. .6333 of consensus"
misc_feature    complement(50811. .51360)
                /note="match: GSS: Em:AQ536659"
repeat_region   52289. .56182
                /note="L1PA3 repeat: matches 2255. .6146 of consensus"
repeat_region   56710. .57178
                /note="MLT1C repeat: matches 1. .463 of consensus"
repeat_region   57746. .57868
                /note="MER33 repeat: matches 202. .323 of consensus"
repeat_region   57869. .58554
                /note="MER44C repeat: matches 1. .727 of consensus"
repeat_region   58555. .58752
                /note="MER33 repeat: matches 1. .202 of consensus"
repeat_region   58890. .59200
                /note="AluSq repeat: matches 1. .307 of consensus"
repeat_region   59630. .60070
```

```
                  /note="L1MB3 repeat: matches 5737. .6184 of consensus"
  misc_feature    59956. .60646
                  /note="match: GSS: Em:B82923"
  repeat_region   61550. .61582
                  /note="L1ME2 repeat: matches 6133. .6164 of consensus"
  repeat_region   61583. .62103

  Query Match             86.8%;  Score 551.4;  DB 9;  Length 128468;
  Best Local Similarity   93.9%;  Pred. No. 8.9e-163;
  Matches  596;  Conservative    0;  Mismatches   36;  Indels    3;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db      33791 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAACTGTAAAACTA 33732

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||  ||||||||||||||| ||||||| |||||||||||||||||||| |  |||||
Db      33731 CAAATCATTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACTGCAGACCC 33672

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              || || |  ||| |||||||||||||| ||||||||||||||||||||||||||||||||
Db      33671 CT-GATCAACCTCCTAGCCCATGCTCCAATGTTAATGACATTGAAGGCACCCCTCCCGAG 33613

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||| ||||||||||| ||||||||||||||| ||
Db      33612 GAAATCTCAACTGCACAACCCCTACTATGACCCAATTCAGCAGGAAGCAGTTAGAGCAGT 33553

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33552 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 33493

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGAAGGTGACT 360
              |||||||||||||||||||||||| || |||||||||||||||||||||||||||||||
Db      33492 AGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGAAGGTGACC 33433

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||| |||| |||||||||||||||| ||||||||||||||| |||||||||||||||.
Db      33432 GCATCTACCTTTAAACATGGGGCTTGCCACTTAGCTCACACCCAACCAATCAGAGAGCTC 33373

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||| | ||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db      33372 ACTAAGAAGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCAATTGCCTG 33313

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||| |||||||||||||||| ||||||||||||||||||| ||| | |||| ||
Db      33312 AGAGCACAGTGGGAGGGACAAGGATCCGGATATAAACCCAGGCATTTGAGACAGCAATGG 33253

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||| ||||||||||||||||||  ||||||||| ||||||||||||||||||||||||||
Db      33252 CAATCCCCTTTGGGTCCCCTCC--TTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 33195

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db      33194 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 33160

RESULT 22
AL139038
LOCUS       AL139038              140756 bp    DNA     linear   PRI 15-JUL-2001
DEFINITION  Human DNA sequence from clone RP11-456B18 on chromosome 13,
            complete sequence.
ACCESSION   AL139038
VERSION     AL139038.18  GI:14800148
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 140756)
  AUTHORS   Dunn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 17, 2001 this sequence version replaced gi:13751261.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-456B18 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-456B18 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-456B18 is at 140756 in this
            sequence. The true left end of clone RP11-78L16 is at 68441 in this
            sequence. The true right end of clone RP11-108H9 is at 100 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .140756
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-456B18"
                     /clone_lib="RPCI-11.2"
     repeat_region   1830. .2326
                     /note="LOR1a repeat: matches 1. .497 of consensus"
     repeat_region   3932. .3965
                     /note="17 copies 2 mer gt 94% conserved"
     repeat_region   4632. .5067
                     /note="LOR1b repeat: matches 1. .461 of consensus"
     repeat_region   5733. .5893
                     /note="L1ME3 repeat: matches 6002. .6146 of consensus"
     repeat_region   6000. .6083
                     /note="MER34 repeat: matches 454. .539 of consensus"
     repeat_region   6108. .6171
                     /note="HERV23 repeat: matches 993. .1056 of consensus"
     repeat_region   7143. .7840
                     /note="LTR8 repeat: matches 1. .691 of consensus"
     repeat_region   8059. .8292
                     /note="LTR16A repeat: matches 203. .442 of consensus"
     repeat_region   8491. .8686
                     /note="LTR29 repeat: matches 6. .203 of consensus"
     repeat_region   8725. .8926
                     /note="MER4D repeat: matches 483. .677 of consensus"
     repeat_region   8992. .9049
                     /note="MER4D repeat: matches 410. .465 of consensus"
     repeat_region   9062. .9281
                     /note="110 copies 2 mer tt 59% conserved"
     repeat_region   9398. .9530
                     /note="MER41C repeat: matches 1. .131 of consensus"
     repeat_region   9623. .9732
                     /note="LTR1 repeat: matches 675. .785 of consensus"
     repeat_region   9970. .10333
                     /note="MLT2D repeat: matches 1. .486 of consensus"
     repeat_region   10400. .10545
                     /note="LTR29 repeat: matches 2. .151 of consensus"
     repeat_region   10549. .10935
                     /note="MER67D repeat: matches 12. .391 of consensus"
     repeat_region   11249. .11332
                     /note="MIR repeat: matches 69. .144 of consensus"
     repeat_region   11923. .12159
                     /note="MER4D repeat: matches 1. .234 of consensus"
     repeat_region   12164. .12767
                     /note="MER4D repeat: matches 396. .973 of consensus"
     repeat_region   12852. .13221
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   13222. .14798
                     /note="THE1B-INTERNAL repeat: matches 1. .1580 of
                     consensus"
     repeat_region   14799. .15173
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   16144. .16346
                     /note="MER54B repeat: matches 585. .789 of consensus"
     repeat_region   16339. .16503
                     /note="MER54B repeat: matches 112. .278 of consensus"
     repeat_region   16504. .16571
                     /note="MER66A repeat: matches 1. .69 of consensus"
     repeat_region   16570. .16629
                     /note="MER41B repeat: matches 283. .341 of consensus"
     repeat_region   16599. .16902
                     /note="MER66A repeat: matches 163. .478 of consensus"
     repeat_region   16903. .17014
                     /note="MER54B repeat: matches 3. .115 of consensus"
     repeat_region   17195. .17491
                     /note="AluY repeat: matches 1. .297 of consensus"
     repeat_region   18120. .19291
                     /note="L1P repeat: matches 1726. .2897 of consensus"
     repeat_region   19460. .19584
                     /note="25 copies 5 mer atttt 62% conserved"
     repeat_region   19686. .19745
                     /note="12 copies 5 mer tttaa 71% conserved"
     repeat_region   20453. .20914
                     /note="L1MD2 repeat: matches 5745. .6199 of consensus"
     repeat_region   20913. .21222
                     /note="L1MD2 repeat: matches 5277. .5594 of consensus"
     repeat_region   21308. .21570
                     /note="LTR9 repeat: matches 357. .625 of consensus"
     repeat_region   21645. .21984
                     /note="LTR19B repeat: matches 55. .393 of consensus"
     repeat_region   22081. .22583
                     /note="L1MD2 repeat: matches 4726. .5242 of consensus"
     repeat_region   23498. .23784
                     /note="AluJb repeat: matches 1. .305 of consensus"
     repeat_region   23873. .24169
                     /note="AluSg repeat: matches 1. .295 of consensus"
     repeat_region   25044. .25238
                     /note="MIR repeat: matches 31. .238 of consensus"
     repeat_region   25404. .25696
                     /note="AluSx repeat: matches 20. .311 of consensus"
     repeat_region   25832. .26218
                     /note="MSTC repeat: matches 1. .405 of consensus"
     repeat_region   26221. .27865
                     /note="MSTC-internal repeat: matches 2. .1651 of consensus"
     repeat_region   27907. .28659
                     /note="L1PB2 repeat: matches 5400. .6155 of consensus"
     repeat_region   28710. .29043
                     /note="MSTD repeat: matches 30. .394 of consensus"
     repeat_region   29634. .29697
                     /note="32 copies 2 mer ta 79% conserved"
     repeat_region   29729. .29752
                     /note="12 copies 2 mer tg 95% conserved"
     repeat_region   30688. .30874
                     /note="MER61A repeat: matches 160. .354 of consensus"
     repeat_region   31648. .31785
                     /note="L1MC4 repeat: matches 6409. .6540 of consensus"
     repeat_region   31813. .32243
                     /note="L1MC4 repeat: matches 6629. .7089 of consensus"
     repeat_region   32449..32821
                     /note="THE1C repeat: matches 1..371 of consensus"
     repeat_region   32824..34464
                     /note="THE1C-internal repeat: matches 3..1651 of
                     consensus"
     repeat_region   34465..34825
                     /note="THE1C repeat: matches 1..360 of consensus"
     repeat_region   34804..35004
                     /note="MER67C repeat: matches 202..403 of consensus"
     repeat_region   39792..39868
                     /note="L1PA12 repeat: matches 6076..6152 of consensus"
     repeat_region   40953..41258
                     /note="AluY repeat: matches 1..308 of consensus"
     repeat_region   41279..41308
                     /note="15 copies 2 mer aa 86% conserved"
     repeat_region   42744..42933
                     /note="L1MA8 repeat: matches 6064..6256 of consensus"
     repeat_region   42973..43275
                     /note="AluSx repeat: matches 1..303 of consensus"
     misc_feature    complement(43560..43991)
                     /note="match: STS: Em:HSPH07E5"
     repeat_region   45196..45584
                     /note="MLT2B repeat: matches 2..399 of consensus"
     repeat_region   45594..45767
                     /note="87 copies 2 mer ta 77% conserved"
     repeat_region   45800..45851
                     /note="MLT2B repeat: matches 394..448 of consensus"
     repeat_region   48118..48487
                     /note="THE1C repeat: matches 1..370 of consensus"
     repeat_region   48708..48806
                     /note="AluJo/FRAM repeat: matches 197..297 of consensus"
     repeat_region   49853..50271
                     /note="L1MEc repeat: matches 2274..2369 of consensus"
     repeat_region   50637..50927
                     /note="L1ME repeat: matches 974..1253 of consensus"
     repeat_region   52250..52349
                     /note="MIR repeat: matches 131..228 of consensus"
     repeat_region   52696..52996
                     /note="AluSg repeat: matches 1..293 of consensus"

  Query Match             86.8%;  Score 551.4;  DB 9;  Length 140756;
  Best Local Similarity   94.3%;  Pred. No. 9e-163;
  Matches  584;  Conservative    0;  Mismatches   31;  Indels    4;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||||||||||||
Db      69181 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGACTCTTCCAGAATTAAAACTGTAAAACTA 69240

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||||||||||||||||||| ||||||| |||||||||||||||||||| | ||||||
Db      69241 CAAATTGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACTGCGGACCC 69300

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db      69301 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCTCCTGAG 69360

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||| |||| |    ||||| ||
Db      69361 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCAGGAAAC----AGAGCAGT 69416

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              |||  |||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      69417 CATTGGCCAACTTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 69476

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||| |||||||||||||||||| |||||||||||||||||||||||||||||||
Db      69477 AGGACTAGTTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 69536

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||||
Db      69537 GCATCCACCTTTAAACATGGGGCTTGCAACTTAACTCACACCCAACCAATCAGAGAGCTC 69596

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||||
Db      69597 ACTAAAATGCTAACTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 69656

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              | ||||||||||||||||||||||||||||||| ||||||| |||||||||||||||| |
Db      69657 AAAGCACAGCGGGAGGGACAAGGATCGGGATATGAACCCAGACATTCGAGCCGGCAACAG 69716

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              |||||||||||||||||||||||||  |||||| ||||||||||||||||||||||||||
Db      69717 CAACCCCCTTTGGGTCCCCTCCCTTCCTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 69776

Qy        601 ATTAAATCTTGCAACTGAA 619
               |||||||||||||||| |
Db      69777 GTTAAATCTTGCAACTGCA 69795


RESULT 23
AC021774/c
LOCUS       AC021774               167366 bp    DNA     linear   HTG 30-MAR-2000
DEFINITION  Homo sapiens clone RP11-13C18, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   AC021774
VERSION     AC021774.3  GI:7341907
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167366)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-13C18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167366)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2000 this sequence version replaced gi:6940059.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3315
                Center clone name: 13_C_18
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161819 bases at least Q40
                Consensus quality: 164760 bases at least Q30
                Consensus quality: 165688 bases at least Q20
                Insert size: 170000; agarose-fp
            Insert size: 166466; sum-of-contigs
            Quality coverage: 4.8 in Q20 bases; agarose-fp
            Quality coverage: 4.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     1272: contig of 1272 bp in length
            *      1273     1372: gap of 100 bp
            *      1373     4088: contig of 2716 bp in length
            *      4089     4188: gap of 100 bp
            *      4189     6500: contig of 2312 bp in length
            *      6501     6600: gap of 100 bp
            *      6601    15485: contig of 8885 bp in length
            *     15486    15585: gap of 100 bp
            *     15586    27637: contig of 12052 bp in length
            *     27638    27737: gap of 100 bp
            *     27738    39958: contig of 12221 bp in length
            *     39959    40058: gap of 100 bp
            *     40059    54618: contig of 14560 bp in length
            *     54619    54718: gap of 100 bp
            *     54719    82725: contig of 28007 bp in length
            *     82726    82825: gap of 100 bp
            *     82826   123232: contig of 40407 bp in length
            *    123233   123332: gap of 100 bp
            *    123333   167366: contig of 44034 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167366
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-13C18"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1272
                     /note="assembly_fragment"
     misc_feature    1373. .4088
                     /note="assembly_fragment"
     misc_feature    4189. .6500
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    6601. .15485
                     /note="assembly_fragment"
     misc_feature    15586. .27637
                     /note="assembly_fragment"
     misc_feature    27738. .39958
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    40059. .54618
                     /note="assembly_fragment"
     misc_feature    54719. .82725
                     /note="assembly_fragment"
     misc_feature    82826. .123232
                     /note="assembly_fragment"
     misc_feature    123333. .167366
                     /note="assembly_fragment"
ORIGIN

  Query Match             86.8%;  Score 551.4;  DB 2;  Length 167366;
  Best Local Similarity   93.9%;  Pred. No. 9.3e-163;
  Matches  596;  Conservative    0;  Mismatches   36;  Indels    3;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db       7535 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAACTGTAAAACTA 7476

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||  ||||||||||||| ||||||| |||||||||||||||||||||| |  |||||
Db       7475 CAAATCATTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACTGCAGACCC 7416

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              || || |  ||| |||||||||||||| ||||||||||||||||||||||||||||||||
Db       7415 CT-GATCAACCTCCTAGCCCATGCTCCAATGTTAATGACATTGAAGGCACCCCTCCCGAG 7357

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||| ||||||||||| ||||||||||||||| ||
Db       7356 GAAATCTCAACTGCACAACCCCTACTATGACCCAATTCAGCAGGAAGCAGTTAGAGCAGT 7297

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7296 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 7237

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||| || |||||||||||||||||||||||||||||| 
Db       7236 AGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 7177

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||| |||| ||||||||||||||||| |||||||||||||||| ||||||||||||||||
Db       7176 GCATCTACCTTTAAACATGGGGCTTGCCACTTAGCTCACACCCAACCAATCAGAGAGCTC 7117

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||| | ||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       7116 ACTAAGAAGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCAATTGCCTG 7057

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||| |||||||||||||||| ||||||||||||||||||||| ||| | |||| ||
Db       7056 AGAGCACAGTGGGAGGGACAAGGATCCGGATATAAACCCAGGCATTTGAGACAGCAATGG 6997

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||| ||||||||||||||||||  ||||||||| ||||||||||||||||||||||||||
Db       6996 CAATCCCCTTTGGGTCCCCTCC--TTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 6939

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       6938 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 6904

RESULT 24
AL607153
LOCUS       AL607153              183499 bp    DNA     linear   HTG 21-OCT-2001
DEFINITION  Homo sapiens chromosome 13 clone RP13-276D12.
ACCESSION   AL607153
VERSION     AL607153.3  GI:15982188
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Lovell,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2001 this sequence version replaced gi:15962724.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bB276D12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 183433 bases at least Q40
            Consensus quality: 183439 bases at least Q30
            Consensus quality: 183441 bases at least Q20
            Insert size: 183499; sum-of-contigs
            Insert size: 179011; 5.6% error; agarose-fp
            Quality coverage: 12.84x in Q20 bases; sum-of-contigs Quality
            coverage: 13.23x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   183499: contig of 183499 bp in length.
FEATURES             Location/Qualifiers
     source          1. .183499
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP13-276D12"
                     /clone_lib="RPCI-13.2"
     misc_feature    1. .183499
                     /note="assembly_fragment:05362"
ORIGIN

  Query Match            86.8%;  Score 551.4;  DB 2;  Length 183499;
  Best Local Similarity  94.3%;  Pred. No. 9.4e-163;
  Matches  584;  Conservative    0;  Mismatches   31;  Indels    4;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||||||||||||
Db      12044 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGACTCTTCCAGAATTAAAACTGTAAAACTA 12103

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||||||||||||||||||| ||||||| |||||||||||||||||||| | ||||||
Db      12104 CAAATTGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACTGCGGACCC 12163

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db      12164 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCTCCTGAG 12223

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||| |||| |    ||||| ||
Db      12224 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCAGGAAAC----AGAGCAGT 12279

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||  |||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      12280 CATTGGCCAACTTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 12339

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||| |||||||||||||||||| |||||||||||||||||||||||||||||||
Db      12340 AGGACTAGTTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 12399

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||||
Db      12400 GCATCCACCTTTAAACATGGGGCTTGCAACTTAACTCACACCCAACCAATCAGAGAGCTC 12459

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||||
Db      12460 ACTAAAATGCTAACTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 12519

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              | ||||||||||||||||||||||||||||||| ||||||| |||||||||||||||| |
Db      12520 AAAGCACAGCGGGAGGGACAAGGATCGGGATATGAACCCAGACATTCGAGCCGGCAACAG 12579

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              |||||||||||||||||||||||||  |||||| ||||||||||||||||||||||||||
Db      12580 CAACCCCCTTTGGGTCCCCTCCCTTCCTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 12639

Qy        601 ATTAAATCTTGCAACTGAA 619
               |||||||||||||||| |
Db      12640 GTTAAATCTTGCAACTGCA 12658
```

```
RESULT 25
AP002790/c
LOCUS       AP002790              164211 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-732P12 map 18q21, WORKING
            DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION   AP002790
VERSION     AP002790.1  GI:9188576
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164211)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 164,211 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 164211)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-732P12
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 154344 bases at least Q40
                Consensus quality: 159150 bases at least Q30
                Consensus quality: 161200 bases at least Q20
                Insert size: 162311; sum-of-contigs
                Quality coverage: 5.05x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            20 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                    1     28606 contig of  28606 bp in length
                28707     55973 contig of  27267 bp in length
                56074     73701 contig of  17628 bp in length
                73802     92513 contig of  18712 bp in length
                92614    102171 contig of   9558 bp in length
               102272    111253 contig of   8982 bp in length
               111354    120167 contig of   8814 bp in length
               120268    127637 contig of   7370 bp in length
               127738    134267 contig of   6530 bp in length
               134368    138746 contig of   4379 bp in length
               138847    143265 contig of   4419 bp in length
               143366    147643 contig of   4278 bp in length
               147744    150342 contig of   2599 bp in length
               150443    153002 contig of   2560 bp in length
               153103    155224 contig of   2122 bp in length
               155325    156567 contig of   1243 bp in length
               156668    158332 contig of   1665 bp in length
               158433    160143 contig of   1711 bp in length
               160244    162721 contig of   2478 bp in length
               162822    164211 contig of   1390 bp in length.
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     28606: contig of 28606 bp in length
*   28607     28706: gap of 100 bp
*   28707     55973: contig of 27267 bp in length
*   55974     56073: gap of 100 bp
*   56074     73701: contig of 17628 bp in length
*   73702     73801: gap of 100 bp
*   73802     92513: contig of 18712 bp in length
*   92514     92613: gap of 100 bp
*   92614    102171: contig of 9558 bp in length
*  102172    102271: gap of 100 bp
*  102272    111253: contig of 8982 bp in length
*  111254    111353: gap of 100 bp
*  111354    120167: contig of 8814 bp in length
*  120168    120267: gap of 100 bp
*  120268    127637: contig of 7370 bp in length
*  127638    127737: gap of 100 bp
*  127738    134267: contig of 6530 bp in length
*  134268    134367: gap of 100 bp
*  134368    138746: contig of 4379 bp in length
*  138747    138846: gap of 100 bp
*  138847    143265: contig of 4419 bp in length
*  143266    143365: gap of 100 bp
*  143366    147643: contig of 4278 bp in length
*  147644    147743: gap of 100 bp
*  147744    150342: contig of 2599 bp in length
*  150343    150442: gap of 100 bp
*  150443    153002: contig of 2560 bp in length
*  153003    153102: gap of 100 bp
*  153103    155224: contig of 2122 bp in length
*  155225    155324: gap of 100 bp
*  155325    156567: contig of 1243 bp in length
*  156568    156667: gap of 100 bp
*  156668    158332: contig of 1665 bp in length
*  158333    158432: gap of 100 bp
*  158433    160143: contig of 1711 bp in length
*  160144    160243: gap of 100 bp
*  160244    162721: contig of 2478 bp in length
*  162722    162821: gap of 100 bp
*  162822    164211: contig of 1390 bp in length.
FEATURES             Location/Qualifiers
     source          1..164211
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q21"
                     /clone="RP11-732P12"
     misc_feature    1..28606
                     /note="assembly_fragment"
     misc_feature    28707..55973
                     /note="assembly_fragment"
     misc_feature    56074..73701
                     /note="assembly_fragment"
     misc_feature    73802..92513
                     /note="assembly_fragment"
     misc_feature    92614..102171
                     /note="assembly_fragment"
     misc_feature    102272..111253
                     /note="assembly_fragment"
     misc_feature    111354..120167
                     /note="assembly_fragment"
     misc_feature    120268..127637
                     /note="assembly_fragment"
     misc_feature    127738..134267
                     /note="assembly_fragment"
```

```
     misc_feature    134368..138746
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    138847..143265
                     /note="assembly_fragment"
     misc_feature    143366..147643
                     /note="assembly_fragment"
     misc_feature    147744..150342
                     /note="assembly_fragment"
     misc_feature    150443..153002
                     /note="assembly_fragment"
     misc_feature    153103..155224
                     /note="assembly_fragment"
     misc_feature    155325..156567
                     /note="assembly_fragment"
     misc_feature    156668..158332
                     /note="assembly_fragment"
     misc_feature    158433..160143
                     /note="assembly_fragment"
     misc_feature    160244..162721
                     /note="assembly_fragment"
     misc_feature    162822..164211
                     /note="assembly_fragment"
ORIGIN

  Query Match             86.6%;  Score 550.2;  DB 2;  Length 164211;
  Best Local Similarity   91.7%;  Pred. No. 2.2e-162;
  Matches  582;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db       6465 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACTA 6406

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||  |||||||||||||| ||||||| ||||||||||||||||| |||   |||||
Db       6405 CAAATCATTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCACAGACCC 6346

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||| ||||||||||||||||||| ||||||||||||  ||||||| ||||||| ||
Db       6345 CTGGACCAGCCTGCTAGCCCATGCTCCAATGTTAATGACAGCGAAGGCAACCCTCCCAAG 6286

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||||||||| |||||| |||||||||||| |||||
Db       6285 GAAATCTCAACTGCACAACCCCTACTATGCCCCAGTTCAGCAGGAAGCAGTTAGGGCGGT 6226

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db       6225 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTCAGAGGGGGTACTGAGAGAC 6166

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||| |||||||||||||||||||
Db       6165 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAACCTAGCTGGGAAGGTGACT 6106

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||| ||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       6105 GCATTCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 6046

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       6045 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 5986

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||| |||||||||||||||| |||||||||||||||||||||||||||||||||
Db       5985 AGAGCACAGTGGGAGGGACAAGGATCAGGATATAAACCCAGGCATTCGAGCCGGCAACGG 5926

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||  |   |
Db       5925 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTAAATCTT 5866

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
                 |   |    ||   ||| |||||||  |||||
Db       5865 TCAACTGCAAAAAATAAAAATAAAAAAATTAAAAA 5831
```

```
RESULT 26
AC064801
LOCUS       AC064801              167843 bp    DNA     linear   PRI 30-JUL-2002
DEFINITION  Homo sapiens chromosome 18, clone RP11-120K19, complete sequence.
ACCESSION   AC064801
VERSION     AC064801.10  GI:22004284
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167843)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-120K19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167843)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 167843)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 167843)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 30, 2002 this sequence version replaced gi:21700673.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8973
                Center clone name: 120_K_19
            --------------.
FEATURES             Location/Qualifiers
     source          1..167843
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-120K19"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(1059..1111)
                     /rpt_family="MIR"
     repeat_region   complement(1112..1424)
                     /rpt_family="AluSx"
     repeat_region   complement(1425..1557)
                     /rpt_family="MIR"
     unsure          1646..1651
                     /note="<30 qual SNGL region"
     unsure          1689..1693
                     /note="<30 qual SNGL region"
     repeat_region   complement(1989..2161)
                     /rpt_family="MER97c"
     repeat_region   2168..2463
                     /rpt_family="AluSx"
     repeat_region   complement(2489..2622)
                     /rpt_family="FLAM_C"
     repeat_region   complement(3544..3967)
                     /rpt_family="MER97a"
     repeat_region   4384..4633
                     /rpt_family="AluSq"
     repeat_region   complement(5110..5257)
                     /rpt_family="MIR3"
     repeat_region   complement(5559..5697)
                     /rpt_family="MER117"
     repeat_region   6353..6769
                     /rpt_family="MLT1C"
     repeat_region   6770..6846
                     /rpt_family="LTR16C"
     repeat_region   complement(6847..6893)
                     /rpt_family="MIR"
     repeat_region   complement(7336..8363)
                     /rpt_family="L1PA3"
     repeat_region   8560..8867
                     /rpt_family="AluSc"
     repeat_region   8962..9273
                     /rpt_family="AluSq"
     repeat_region   11717..11744
                     /rpt_family="(TAAAA)n"
     repeat_region   12041..12266
                     /rpt_family="MER50"
     repeat_region   12267. .12511
                     /rpt_family="AluSc"
     repeat_region   12512. .13031
                     /rpt_family="MER50"
     repeat_region   complement(14240. .14530)
                     /rpt_family="AluJo"
     repeat_region   14881. .15288
                     /rpt_family="LTR16A1"
     repeat_region   15290. .15645
                     /rpt_family="THE1C"
     repeat_region   complement(15654. .15794)
                     /rpt_family="MER5A"
     repeat_region   15935. .16489
                     /rpt_family="MLT2A2"
     repeat_region   complement(17190. .17503)
                     /rpt_family="AluSx"
     repeat_region   complement(17524. .17861)
                     /rpt_family="L1MEd"
     repeat_region   complement(17998. .18430)
                     /rpt_family="MER65A"
     repeat_region   19847. .19874
                     /rpt_family="AT_rich"
     repeat_region   complement(19903. .20222)
                     /rpt_family="AluJo"
     repeat_region   complement(21102. .21388)
                     /rpt_family="AluJb"
     repeat_region   21410. .21692
                     /rpt_family="L2"
     repeat_region   21693. .21985
                     /rpt_family="AluSx"
     repeat_region   21986. .22149
                     /rpt_family="L2"
     repeat_region   22201. .22391
                     /rpt_family="LTR67"
     repeat_region   22657. .22934
                     /rpt_family="AluJb"
     repeat_region   22940. .23062
                     /rpt_family="L1ME3"
     repeat_region   24066. .24354
                     /rpt_family="L1ME4A"
     repeat_region   25031. .25348
                     /rpt_family="AluSx"
     repeat_region   25583. .26009
                     /rpt_family="MLT1C"
     repeat_region   28144. .28456
                     /rpt_family="AluY"
     repeat_region   complement(28729. .28847)
                     /rpt_family="L1MC/D"
     repeat_region   29217. .29290
                     /rpt_family="L2"
     repeat_region   29341. .29442
                     /rpt_family="MLT1I"
     repeat_region   29557. .29579
                     /rpt_family="(GAAAA)n"
     repeat_region   complement(29806. .30079)
                     /rpt_family="AluJb"

  Query Match             86.6%;  Score 550.2;  DB 9;  Length 167843;
  Best Local Similarity   91.7%;  Pred. No. 2.2e-162;
  Matches  582;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      50488 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACTA 50547

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||  ||||||||||||||| ||||||| ||||||||||||||||| |||   |||||
Db      50548 CAAATCATTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCACAGACCC 50607

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||| ||||||||||||||||||| ||||||||||||  ||||||| ||||||| ||
Db      50608 CTGGACCAGCCTGCTAGCCCATGCTCCAATGTTAATGACAGCGAAGGCAACCCTCCCAAG 50667
```

```
Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||||||||| |||||| |||||||||||| |||||
Db      50668 GAAATCTCAACTGCACAACCCCTACTATGCCCCAGTTCAGCAGGAAGCAGTTAGGGCGGT 50727

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db      50728 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTCAGAGGGGGTACTGAGAGAC 50787

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||| |||||||||||||||||||
Db      50788 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAACCTAGCTGGGAAGGTGACT 50847

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||| ||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db      50848 GCATTCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 50907

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      50908 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 50967

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||| |||||||||||||||| |||||||||||||||||||||||||||||||||
Db      50968 AGAGCACAGTGGGAGGGACAAGGATCAGGATATAAACCCAGGCATTCGAGCCGGCAACGG 51027

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||  |   |
Db      51028 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTAAATCTT 51087

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
                 |   |    ||   ||| |||||||  |||||
Db      51088 TCAACTGCAAAAAATAAAAATAAAAAAATTAAAAA 51122

RESULT 27
CNS01DSH
LOCUS       CNS01DSH              180523 bp    DNA     linear   PRI 20-AUG-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-305I3 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL121784
VERSION     AL121784.5  GI:15282089
KEYWORDS    HTG; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180523)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180523)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Aug 23, 2001 this sequence version replaced gi:12001714.
            --------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            ---------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : C-2547L24 (AC=AL135818)
            Downstream BAC (overlapping the SP6 end) : R-580E16 ---------------
            Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.78x in Q20 bases; sum-of-contigs
            ---------------
            ----------------------
            Overall quality chart :
            Range    :    bases
            0        :
            1  -  9  :
            10 - 19  :       18
            20 - 29  :       87
            30 - 39  :      360
            40 - 49  :     3700
            50 - 59  :    10918
            60 - 69  :    13753
            70 - 79  :    29245
            80 - 89  :    58279
            90 - 99  :    64163
            -----------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES             Location/Qualifiers
     source          1. .180523
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-305I3"
                     /clone_lib="RPCI-11"
     STS             1561. .1825
                     /note="matching EMBL:R74544
                     RHdb:RH53560
                     dbSTS:STS42868
                     Identified using the e-PCR software (G. Schuler)"
     STS             69085. .69312
                     /note="matching EMBL:G03465
                     RHdb:RH53501
                     RHdb:RH3654
                     dbSTS:STS1070
                     Identified using the e-PCR software (G. Schuler)"
     STS             71190. .71397
                     /note="matching EMBL:AA010373
                     RHdb:RH75482
                     dbSTS:STS52570
                     Identified using the e-PCR software (G. Schuler)"
ORIGIN

  Query Match             86.5%;  Score 549;  DB 9;  Length 180523;
  Best Local Similarity   92.9%;  Pred. No. 5.4e-162;
  Matches  590;  Conservative    0;  Mismatches   35;  Indels   10;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||| ||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db        537 CCCTTTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACTA 596

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||||||||||| ||| |||||||
Db        597 CAAATAGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCATGGACCC 656

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db        657 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTGATGACATCGAAGGCACCCCTCCCGAG 716

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||| ||
Db        717 GAAATCTCAACTGCACAACCCCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 776

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
                |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db        777 TGTCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGTACTGAGAGAC 836

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        837 AGGACTAGCTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 896

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| ||||||  |||||||||||||||||||||||||||||||||||||||||
Db        897 ACATCCACCTTTAAACACAGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 956
```

```
Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||| |||||
Db        957 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATCGCCTG 1016

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||||||||||||||| |||| ||
Db       1017 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCATTCGAGCCAGCAATGG 1076

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              | |||| |||||||||||||||||||||||||| |||| ||          |||||||||
Db       1077 CTACCCTCTTTGGGTCCCCTCCCTTTGTATGGAGCTCAGT----------TTTCACTCT 1126

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| ||||| ||| ||| |||
Db       1127 ATTAAATCTTGCAACTGCAAAAACAAACAAACAAA 1161


RESULT 28
AC007374
LOCUS       AC007374              190565 bp    DNA     linear   PRI 27-MAR-2001
DEFINITION  Homo sapiens chromosome 14 clone RP11-325L17 map 14q31, complete
            sequence.
ACCESSION   AC007374
VERSION     AC007374.6  GI:13450000
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190565)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190565)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dickhoff,R.,
            James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S., Ratcliffe,A.,
            Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 190565)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
COMMENT     On Mar 27, 2001 this sequence version replaced gi:8247780.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399

            ------------.
FEATURES             Location/Qualifiers
     source          1. .190565
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q31"
                     /clone="RP11-325L17"
                     /clone_lib="RPCI human BAC library 11"
     unsure          985. .1035
```

```
                    /note="low quality data"
     unsure         3865. .3870
                    /note="low quality data"
     unsure         14875. .14880
                    /note="low quality data"
     unsure         17060. .17095
                    /note="low quality data"
     unsure         17488. .17535
                    /note="low quality data"
     unsure         19433
                    /note="low quality data"
     unsure         46474. .46580
                    /note="low quality data"
     unsure         46695. .46700
                    /note="low quality data"
     unsure         55163. .55242
                    /note="low quality data"
     unsure         110860. .110875
                    /note="low quality data"
     unsure         113613. .113617
                    /note="low quality data"
     misc_feature   116230. .116782
                    /note="Single clone coverage. PCR and transposons tried."
     unsure         117677. .117680
                    /note="low quality data"
     unsure         117730. .117748
                    /note="low quality data"
     unsure         118505. .118736
                    /note="low quality data"
     unsure         144847. .144850
                    /note="low quality data"
     unsure         152067. .152071
                    /note="low quality data"
     unsure         158953. .158990
                    /note="low quality data"
     unsure         177622. .177705
                    /note="low quality data"
ORIGIN

  Query Match             86.5%;  Score 549;  DB 9;  Length 190565;
  Best Local Similarity   92.9%;  Pred. No. 5.5e-162;
  Matches  590;  Conservative    0;  Mismatches   35;  Indels   10;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||| ||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      23017 CCCTTTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACTA 23076

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||||||||||| ||| |||||||
Db      23077 CAAATAGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCATGGACCC 23136

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db      23137 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTGATGACATCGAAGGCACCCCTCCCGAG 23196

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||| ||
Db      23197 GAAATCTCAACTGCACAACCCCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 23256

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      23257 TGTCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAGAGAC 23316

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      23317 AGGACTAGCTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 23376

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| ||||||  |||||||||||||||||||||||||||||||||||||||||
Db      23377 ACATCCACCTTTAAACACAGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 23436

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||| |||||
Db      23437 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATCGCCTG 23496

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||||||||||||||| |||| ||
Db      23497 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCATTCGAGCCAGCAATGG 23556

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              | |||| |||||||||||||||||||||||||| |||| ||          |||||||||
Db      23557 CTACCCTCTTTGGGTCCCCTCCCTTTGTATGGAGCTCAGT----------TTTCACTCT 23606

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| ||||| ||| ||| |||
Db      23607 ATTAAATCTTGCAACTGCAAAAACAAACAAACAAA 23641


RESULT 29
CNS01DVH/c
LOCUS       CNS01DVH               203777 bp    DNA     linear   PRI 04-MAY-2001
DEFINITION  Human chromosome 14 DNA sequence BAC C-2547L24 of library CalTech-D
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL135818
VERSION     AL135818.3  GI:8217905
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 203777)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203777)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Jun 3, 2000 this sequence version replaced gi:7406569.
            -------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            --------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-305I3
            Downstream BAC (overlapping the SP6 end) : R-895M11 (AC=AL133153)
            -------------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.44x in Q20 bases; sum-of-contigs
            --------------
            -----------------------
            Overall quality chart :
            Range     :   bases
            0         :       2
            1  -  9   :      29
            10 -  19  :     188
            20 -  29  :     490
            30 -  39  :    1165
            40 -  49  :    5179
            50 -  59  :    5946
            60 -  69  :    9708
            70 -  79  :   23956
            80 -  89  :   61022
            90 -  99  :   96092
            -----------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES             Location/Qualifiers
     source          1. .203777
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="C-2547L24"
                     /clone_lib="CalTech-D"
     STS             9045..9309
                     /note="matching EMBL:R74544
                     RHdb:RH53560
                     dbSTS:STS42868
                     Identified using the e-PCR software (G. Schuler)"
     STS             16628..16834
                     /note="matching EMBL:G20331
                     RHdb:RH16802
                     dbSTS:STS13937
                     Identified using the e-PCR software (G. Schuler)"
     STS             42379..42560
                     /note="matching EMBL:N80119
                     RHdb:RH46446
                     dbSTS:STS39514
                     Identified using the e-PCR software (G. Schuler)"
     STS             145247..145418
                     /note="matching EMBL:T02957
                     RHdb:RH53817
                     RHdb:RH22284
                     dbSTS:STS19329
                     Identified using the e-PCR software (G. Schuler)"
     STS             162745..163039
                     /note="matching EMBL:R51241
                     RHdb:RH53819
                     dbSTS:STS42905
                     Identified using the e-PCR software (G. Schuler)"
     STS             167890..168003
                     /note="matching EMBL:G38628
                     RHdb:RH97597
                     RHdb:RH83840
                     dbSTS:STS59195
                     Identified using the e-PCR software (G. Schuler)"
ORIGIN

  Query Match             86.5%;  Score 549;  DB 9;  Length 203777;
  Best Local Similarity   92.9%;  Pred. No. 5.5e-162;
  Matches  590;  Conservative    0;  Mismatches   35;  Indels   10;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||| ||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      10333 CCCTTTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACTA 10274

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||||||||||| ||| |||||||
Db      10273 CAAATAGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCATGGACCC 10214

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db      10213 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTGATGACATCGAAGGCACCCCTCCCGAG 10154

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||| ||
Db      10153 GAAATCTCAACTGCACAACCCCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 10094

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      10093 TGTCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAGAGAC 10034

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||| 
Db      10033 AGGACTAGCTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 9974

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| ||||||  |||||||||||||||||||||||||||||||||||||||||
Db       9973 ACATCCACCTTTAAACACAGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 9914

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||| |||||
```

```
Db       9913 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATCGCCTG 9854

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||||||||||||||| |||| ||
Db       9853 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCATTCGAGCCAGCAATGG 9794

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              | |||| |||||||||||||||||||||||||| |||| ||          |||||||||
Db       9793 CTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCAGT----------TTTCACTCT 9744

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| ||||| ||| ||| |||
Db       9743 ATTAAATCTTGCAACTGCAAAAACAAACAAACAAA 9709


RESULT 30
AC109992
LOCUS       AC109992              112544 bp    DNA     linear   PRI 25-OCT-2002
DEFINITION  Homo sapiens 3 BAC RP11-383G6 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC109992
VERSION     AC109992.6  GI:21591818
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112544)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 112544)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 112544)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 112544)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 112544)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 26, 2002 this sequence version replaced gi:20976483.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
            ml.
FEATURES             Location/Qualifiers
     source          1..112544
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-383G6"
     repeat_region   complement(46..1498)
                     /rpt_family="L1M4"
     repeat_region   complement(1499..1804)
                     /rpt_family="AluSc"
     repeat_region   complement(1805..2516)
                     /rpt_family="L1M4"
```

```
     repeat_region   complement(2556..2621)
                     /rpt_family="AluJ/FLAM"
     repeat_region   complement(2622..2786)
                     /rpt_family="L1M4"
     repeat_region   complement(2792..2833)
                     /rpt_family="L1M4"
     repeat_region   complement(2834..3134)
                     /rpt_family="AluY"
     repeat_region   complement(3135..3392)
                     /rpt_family="L1M4"
     repeat_region   5391..5415
                     /rpt_family="AT_rich"
     repeat_region   complement(5509..5805)
                     /rpt_family="AluY"
     repeat_region   6389..6551
                     /rpt_family="MIR"
     repeat_region   6981..7109
                     /rpt_family="AluSg/x"
     repeat_region   7173..7203
                     /rpt_family="(TTTA)n"
     repeat_region   complement(7206..7491)
                     /rpt_family="AluSx"
     repeat_region   complement(7505..7618)
                     /rpt_family="L1MC5"
     repeat_region   7761..8048
                     /rpt_family="AluSx"
     repeat_region   10539..10837
                     /rpt_family="AluSg"
     repeat_region   11017..11061
                     /rpt_family="MADE1"
     repeat_region   11176..11213
                     /rpt_family="AT_rich"
     repeat_region   complement(11215..11374)
                     /rpt_family="U1"
     repeat_region   complement(11387..11681)
                     /rpt_family="AluSx"
     repeat_region   12173..12203
                     /rpt_family="AT_rich"
     repeat_region   12313..12350
                     /rpt_family="A-rich"
     repeat_region   13076..13361
                     /rpt_family="AluSx"
     repeat_region   14236..14529
                     /rpt_family="L1M4"
     repeat_region   complement(15115..15406)
                     /rpt_family="AluSx"
     repeat_region   15454..15878
                     /rpt_family="L1MC5"
     repeat_region   16207..16227
                     /rpt_family="AT_rich"
     repeat_region   16946..17002
                     /rpt_family="AT_rich"
     repeat_region   18059..18361
                     /rpt_family="AluJo"
     repeat_region   19122..19425
                     /rpt_family="AluSx"
     repeat_region   complement(19530..20155)
                     /rpt_family="L2"
     repeat_region   complement(20156..20458)
                     /rpt_family="AluSg1"
     repeat_region   complement(20459..20559)
                     /rpt_family="L2"
     repeat_region   20687..20790
                     /rpt_family="MLT1J2"
     repeat_region   21217..21344
                     /rpt_family="LTR33A"
     repeat_region   complement(21391..21514)
                     /rpt_family="MER5A"
     repeat_region   22051..22099
                     /rpt_family="L2"
     repeat_region   22410..22545

  Query Match          86.3%;  Score 547.8;  DB 9;  Length 112544;
  Best Local Similarity   92.4%;  Pred. No. 1.2e-161;
  Matches  587;  Conservative    0;  Mismatches   47;  Indels    1;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||| || |||||||||||
Db      34048 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 34107

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||||||||||| |||| ||||||
Db      34108 CAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCGCGGACCC 34167

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||| |||||||||| | ||||||||||||||  |||||||||||||||||
Db      34168 CTGGACCGGCCTGTTAGCCCATGCGCTGATGTTAATGACATCAAAGGCACCCCTCCCGAG 34227

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||| |||||||||| |  |||||||||| ||||||||||||||| ||
Db      34228 GAAATCTCAACTGCACCACCCCTACTACGTGCCAATTCAGCAGGAAGCAGTTAGAGCAGT 34287

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||| |||||||||||||||||||||| |||||||||| |||||| ||||||||||||||
Db      34288 CATCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGCTGAGAGTGGGGACTGAGAGAC 34347

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||  |||||||||||| || |||||||||||||||||||||||||||||||
Db      34348 AGGACTAGCTACATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 34407

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              |||||||||| |||||| |||||||||||||||||||||||| || ||||||||||||||
Db      34408 GCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGATCAATCAGAGAGCTC 34467

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||| |||| |||||||||||||||||||||||||||| ||||||
Db      34468 ACTAAAATGCTAATTAGGC-AAAACAGGAGGTAAAGAAATAGCCAATCATCTACTGCCTG 34526

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              | ||||||| ||||||||||||||||||||| |||||||| |||||| |||||||||| |
Db      34527 ACAGCACAGTGGGAGGGACAAGGATCGGGATGTAAACCCACGCATTCAAGCCGGCAACAG 34586

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              |||||||||| ||||||||||| |||||||||| |||||||||||||||||| |||||||
Db      34587 CAACCCCCTTCGGGTCCCCTCCTTTTGTATGGGAGCTCTGTTTTCACTCTATGTCACTCT 34646

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| ||||| |||.||| |||
Db      34647 ATTAAATCTTGCAACTGCAAAAACAAACAAACAAA 34681

Search completed: January 22, 2005, 16:13:43
Job time : 2674.68 secs
```

```
                         GenCore version 5.1.6
                Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 11:11:14 ; Search time 291.418 Seconds
                                              (without alignments)
                                              11438.492 Million cell updates/sec

Title:          US-09-319-156B-6
Perfect score:  635
Sequence:       1 ccctgtatctttaacctcct..........tgaaaaaaaaaaaaaaaaaa 635

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters:      8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_23Sep04:*
                1:  geneseqn1980s:*
                2:  geneseqn1990s:*
                3:  geneseqn2000s:*
                4:  geneseqn2001as:*
                5:  geneseqn2001bs:*
                6:  geneseqn2002as:*
                7:  geneseqn2002bs:*
                8:  geneseqn2003as:*
                9:  geneseqn2003bs:*
                10:  geneseqn2003cs:*
                11:  geneseqn2003ds:*
                12:  geneseqn2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                    %
Result              Query
   No.    Score     Match Length DB  ID                          Description
-----------------------------------------------------------------------------
     1      635   100.0    635  2  AAV43215                    Aav43215 Multiple
     2      635   100.0    635  2  AAX29702                    Aax29702 Clone CL6
     3      635   100.0    635  12  ADG14843                    Adg14843 MSRV asso
     4      626    98.6    635  9  ADB84397                    Adb84397 MSRV-1 as
     5    617.4    97.2   2030  3  AAA63826                    Aaa63826 Nucleotid
     6    546.4    86.0   2074  6  AAD41225                    Aad41225 Human EMB
     7      545    85.8   1329  2  AAV43219                    Aav43219 Multiple
     8      545    85.8   1329  2  AAX29704                    Aax29704 Clone 5M6
     9      545    85.8   1329  12  ADG14849                    Adg14849 MSRV asso
    10    540.2    85.1   2046  8  ACC46747                    Acc46747 Human dit
    11      535    84.3   1329  9  ADB84403                    Adb84403 MSRV-1 as
    12    533.8    84.1   2946  2  AAX77526                    Aax77526 Human sec
    13    533.8    84.1   2946  3  AAZ59468                    Aaz59468 Human sec
    14    533.8    84.1   2946  10  ADC38776                    Adc38776 Human cDN
c   15    532.8    83.9 326014  6  ABK89296                    Abk89296 Human gen
c   16    532.8    83.9 326014  12  ADQ94981                    Adq94981 Human kin
    17    532.2    83.8   2781  5  AAF55630                    Aaf55630 Nucleotid
    18    525.8    82.8   1136  2  AAX25660                    Aax25660 Human end
    19    525.8    82.8   1136  3  AAA59210                    Aaa59210 3' pol ge
    20    522.6    82.3   2782  2  AAX25661                    Aax25661 Human end
    21    522.6    82.3   2782  3  AAA59211                    Aaa59211 5' non co
    22    522.6    82.3   2782  5  AAH20069                    Aah20069 HERV-W en
    23    517.4    81.5   2930  6  AAD24195                    Aad24195 Human syn
    24    516.2    81.3  56093  6  ABL61744                    Abl61744 Colon ade
    25    511.4    80.5   7582  2  AAX25665                    Aax25665 Complete
    26    511.4    80.5   7582  3  AAA59215                    Aaa59215 Human end
    27    510.6    80.4  46340  3  ABN97978                    Abn97978 Human ret
    28    505.8    79.7  10499  3  ABN97929                    Abn97929 Human ret
    29    499.6    78.7   1894  4  AAI14608                    Aai14608 Probe #45
    30    499.6    78.7   1894  4  ABA56337                    Aba56337 Human foe
    31    499.6    78.7   1894  4  AAI35980                    Aai35980 Probe #46
    32    499.6    78.7   1894  4  ABA45822                    Aba45822 Human bre
    33    499.6    78.7   1894  4  ABA25978                    Aba25978 Probe #44
    34    499.6    78.7   1894  4  AAK30018                    Aak30018 Human bon
    35    499.6    78.7   1894  4  AAK04516                    Aak04516 Human bra
    36    499.6    78.7   1894  4  ABS29670                    Abs29670 Human liv
    37    499.6    78.7   1894  5  AAI04422                    Aai04422 Probe #44
    38    499.6    78.7   1894  6  ABS04589                    Abs04589 Human gen
    39    495.6    78.0   3372  2  AAX25663                    Aax25663 Human end
    40    495.6    78.0   3372  3  AAA59213                    Aaa59213 Partial p
    41    493.2    77.7   2052  4  ABA08902                    Aba08902 Human rec
    42    482.6    76.0  22436  9  ADA02882                    Ada02882 Human PAP
    43    482.6    76.0  22436  10  ADB72620                    Adb72620 Human PAP
    44    482.6    76.0  22436  10  ADC85361                    Adc85361 Mouse Pap
    45    482.6    76.0  22436  12  ADM74477                    Adm74477 Human car


                                ALIGNMENTS


RESULT 1
AAV43215
ID   AAV43215 standard; cDNA; 635 BP.
XX
AC   AAV43215;
XX
DT   29-DEC-1998  (first entry)
XX
DE   Multiple sclerosis associated retrovirus fragment 4.
XX
KW   Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW   env gene; rheumatoid arthritis-associated virus; ss.
XX
OS   Multiple sclerosis associated retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..234
FT                   /*tag=  a
FT                   /product= "Encodes protein AAW71067"
XX
PN   WO9823755-A1.
XX
PD   04-JUN-1998.
XX
PF   26-NOV-1997;   97WO-IB001482.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B;
XX
DR   WPI; 1998-322732/28.
DR   P-PSDB; AAW71067.
XX
PT   New nucleic acid from retroviruses - useful for diagnosis, prevention and
PT   treatment of, e.g. multiple sclerosis.
XX
PS   Disclosure; Page 183; 286pp; English.
XX
CC   The present sequence represents a multiple sclerosis (MS) associated
CC   retrovirus (MSRV) genomic fragment used in the method of the invention.
CC   The invention provides complete or partial genomic sequences of the MSRV-
CC   1 pol gene, gag gene and env gene, and polypeptides encoded by these
CC   genes. The invention also provides antibodies raised against the
CC   polypeptides. The genomic sequences, polypeptides and antibodies are also
CC   claimed useful for diagnosing infection by MS and rheumatoid arthritis-
CC   associated viruses, and also for prevention and treatment of infection
CC   with these viruses
XX
SQ   Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 635;  DB 2;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-187;
  Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635


RESULT 2
AAX29702
ID   AAX29702 standard; DNA; 635 BP.
XX
AC   AAX29702;
XX
DT   17-OCT-2003  (revised)
DT   27-AUG-2003  (revised)
DT   08-JUN-1999  (first entry)
XX
DE   Clone CL6-3' from MSRV-1.
XX
KW   Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KW   rheumatoid polyarthritis; ss.
XX
OS   Viruses.
XX
PN   FR2765588-A1.
XX
PD   08-JAN-1999.
XX
PF   07-JUL-1997;   97FR-00008816.
XX
PR   07-JUL-1997;   97FR-00008816.
XX
PA   (INMR ) BIO MERIEUX.
XX
DR   WPI; 1999-098275/09.
DR   P-PSDB; AAW99552.
XX
PT   Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT   multiple sclerosis or rheumatoid polyarthritis.
XX
PS   Claim 1; Page 36-37; 83pp; French.
XX
CC   This sequence represents clone CL6-3' from a novel multiple sclerosis
CC   related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC   prophylactic or therapeutic compositions to inhibit expression of a
CC   multiple sclerosis related virus and/or virus associated with rheumatoid
CC   polyarthritis. (Updated on 27-AUG-2003 to correct OS field.) (Updated on
CC   17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 635;  DB 2;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-187;
  Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
```

```
Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635


RESULT 3
ADG14843
ID   ADG14843 standard; cDNA; 635 BP.
XX
AC   ADG14843;
XX
DT   26-FEB-2004  (first entry)
XX
DE   MSRV associated cDNA #7.
XX
KW   ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX
OS   Multiple sclerosis associated retrovirus.
XX
PN   US2003198647-A1.
XX
PD   23-OCT-2003.
XX
PF   03-APR-2002; 2002US-00114104.
XX
PR   26-NOV-1996;   96US-00756429.
PR   26-NOV-1997;   97US-00979847.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2004-032461/03.
XX
PT   New isolated nucleic acid and their fragments having the pol gene of a
PT   retrovirus, useful for diagnosing, preventing and/or treating multiple
PT   sclerosis and/or rheumatoid arthritis.
XX
PS   Disclosure; SEQ ID NO 102; 193pp; English.
XX
CC   The invention relates to an isolated nucleic acid which comprises the pol
CC   gene of a retrovirus associated with multiple sclerosis or rheumatoid
CC   arthritis. The methods and compositions of the present invention are
CC   useful for diagnosing, preventing and/or treating multiple sclerosis
CC   and/or rheumatoid arthritis. The present sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 635;  DB 12;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-187;
  Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635


RESULT 4
ADB84397
ID   ADB84397 standard; DNA; 635 BP.
XX
AC   ADB84397;
XX
DT   04-DEC-2003  (first entry)
XX
DE   MSRV-1 associated DNA sequence #17.
XX
KW   ds; multiple sclerosis; rheumatoid arthritis; gag; pol;
KW   reverse transcriptase; ribonuclease H.
XX
OS   Unidentified;.
XX
PN   US2003039664-A1.
XX
PD   27-FEB-2003.
XX
PF   26-NOV-1997;   97US-00979847.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (PERR/) PERRON H.
PA   (BESE/) BESEME F.
PA   (BEDI/) BEDIN F.
PA   (PARA/) PARANHOS-BACCALA G.
PA   (KOMU/) KOMURIAN-PRADEL F.
PA   (JOLI/) JOLIVET-REYNAUD C.
PA   (MAND/) MANDRAND B.
PA   (GARS/) GARSON J A.
PA   (TUKE/) TUKE P W.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2003-512253/48.
XX
PT   New isolated or purified nucleic acid associated with multiple sclerosis
PT   and/or rheumatoid arthritis, useful for detecting a virus associated with
PT   multiple sclerosis or rheumatoid arthritis in a biological sample.
XX
PS   Claim 31; Page 76-77; 193pp; English.
XX
CC   The invention relates to an isolated or purified nucleic acid from a
CC   virus associated with multiple sclerosis and/or rheumatoid arthritis,
CC   multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
CC   pol, gag or reverse transcriptase genes (or their fragments) encoding the
CC   proteins or defined peptides (including immunodominant peptides,
CC   antigenic peptides or conserved motifs).Also included are a process for
CC   detecting a virus associated with multiple sclerosis or rheumatoid
CC   arthritis in a biological sample, a nucleic acid probe for the detection
CC   of a virus associated with multiple sclerosis or rheumatoid arthritis, a
CC   primer for the amplification by polymerisation of a nucleic acid of a
CC   viral material associated with multiple sclerosis or rheumatoid
CC   arthritis, a polypeptide exhibiting an inhibitory activity on the
CC   proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC   and an antibody directed against the MSRV-1 virus obtained by
CC   immunologically reacting a human or animal body or cells with an
CC   immunogenic agent consisting of the antigenic polypeptide defined above.
CC   The nucleic acids are useful for detecting a biological sample a virus
CC   associated with multiple sclerosis or rheumatoid arthritis, or for
CC   detecting in a biological sample, the presence of or exposure to a virus
CC   associated with multiple sclerosis or rheumatoid arthritis. The present
CC   sequence is a claimed MSRV-associated sequence whose identity cannot be
CC   accurately determined. Note: The SEQ ID numbers for the sequences as
CC   displayed in the main body of the patent do not match the SEQ ID numbers
CC   in the sequence listing. Consequently those sequences mentioned in the
CC   claims may not be the sequences the authors intended to claim.
XX
SQ   Sequence 635 BP; 188 A; 167 C; 133 G; 138 T; 0 U; 9 Other;

  Query Match             98.6%;  Score 626;  DB 9;  Length 635;
  Best Local Similarity   98.6%;  Pred. No. 5.7e-184;
  Matches  626;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCN 120

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAN 180

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGN 240

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAN 300

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACN 360

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTN 420

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTN 480

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGN 540

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCN 600

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
```

```
RESULT 5
AAA63826
ID   AAA63826 standard; DNA; 2030 BP.
XX
AC   AAA63826;
XX
DT   06-AUG-2003  (revised)
DT   04-DEC-2000  (first entry)
XX
DE   Nucleotide sequence of the MSRV-1 3' env and LTR regions.
XX
KW   MSRV-1; pol region; long terminal repeat; LTR; RU5 region; retrovirus;
KW   ss.
XX
OS   Multiple sclerosis associated retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1629
FT                   /*tag=  a
FT                   /note= "Contains one termination codon"
FT   sig_peptide     1. .81
FT                   /*tag=  b
FT   CAAT_signal     1800. .1807
FT                   /*tag=  c
FT   CAAT_signal     1858. .1864
FT                   /*tag=  d
FT   TATA_signal     1906. .1911
FT                   /*tag=  e
FT   polyA_signal    1996. .2002
FT                   /*tag=  f
XX
PN   WO200047745-A1.
XX
PD   17-AUG-2000.
XX
PF   15-FEB-2000; 2000WO-IB000159.
XX
PR   15-FEB-1999;   99EP-00420041.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Perron H,  Komurian-Pradel F;
XX
DR   WPI; 2000-506097/45.
DR   P-PSDB; AAB08195.
XX
PT   Nucleotide fragment of LTR-RU5 region from Multiple Sclerosis retrovirus
PT   (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological
PT   sample.
XX
PS   Disclosure; Fig 2; 23pp; English.
XX
CC   The present sequence represents the nucleotide sequence corresponding to
CC   the 3' env region and long terminal repeat sequences from clone CL6 of
CC   Multiple Sclerosis retrovirus (MSRV-1). The specification describes a
CC   long terminal repeat (LTR)-RU5 region which encodes the expression of a
CC   MSRV-1 protein. This is unusual for LTRs, in particular in the RU5
CC   region. The sequence includes CAAT and TATA signals which are present in
CC   the U3 and R regions and are not directed towards the CDS indicated in
CC   the features table. Probes and antibodies to the MSRV-1 retrovirus
CC   protein and encoding polynucleotide sequences are used to detect the
CC   presence of MSRV-1 retrovirus in a biological sample. (Updated on 06-AUG-
CC   2003 to correct OS field.)
XX
SQ   Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 U; 0 Other;

  Query Match             97.2%;  Score 617.4;  DB 3;  Length 2030;
  Best Local Similarity   98.3%;  Pred. No. 4.8e-181;
  Matches  624;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||  || ||||||| |||
Db       1396 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1455

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || ||  | |||||||| | |||||||||||||||||||||||||||||||||||||
Db       1456 CAGATGGTCTTACAAATGGAACCCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 1515

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1516 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 1575

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1576 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 1635

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1636 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 1695

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1696 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 1755

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1756 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 1815

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1816 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 1875

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1876 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 1935

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1936 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 1995

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db       1996 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 2030


RESULT 6
AAD41225
ID   AAD41225 standard; cDNA; 2074 BP.
XX
AC   AAD41225;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Human EMBRY-2 cDNA.
XX
KW   Human; embryogenesis associated protein; AIDS; reproductive disorder;
KW   infertility; endometriosis; endometrial tumour; inflammatory disorder;
KW   autoimmune disorder; acquired immune deficiency syndrome; transgenic;
KW   ovarian tumour; contact dermatitis; placenta disorder; preeclampsia;
KW   EMBRY-2; allergy; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             64. .1692
FT                   /*tag=  a
FT                   /product= "EMBRY-2 protein"
FT   sig_peptide     64. .123
FT                   /*tag=  b
FT   mat_peptide     124. .1689
FT                   /*tag=  c
FT                   /product= "Mature EMBRY-2 protein"
XX
PN   WO200248362-A2.
XX
```

```
PD   20-JUN-2002.
XX
PF   14-NOV-2001; 2001WO-US043956.
XX
PR   15-NOV-2000; 2000US-0249407P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Ramkumar J,  Arvizu C;
XX
DR   WPI; 2002-537629/57.
DR   P-PSDB; AAE25054.
XX
PT   New polypeptides of human embryogenesis associated proteins for screening
PT   modulators useful for treating or preventing disorders e.g.
PT   endometriosis, infertility, allergy, preeclampsia.
XX
PS   Claim 59; Page 96-97; 97pp; English.
XX
CC   The invention relates to human embryogenesis associated proteins (EMBRY)
CC   and nucleic acid molecules encoding such proteins. EMBRY sequences are
CC   useful for screening modulators useful for treating or preventing
CC   disorders associated with abnormal expression of EMBRY. The disorders
CC   treated include reproductive disorders such as infertility,
CC   endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory
CC   disorder such as acquired immune deficiency syndrome (AIDS), allergies,
CC   contact dermatitis; disorders of the placenta such as preeclampsia,
CC   abruptio placentae etc. Sequences of the invention are also useful for
CC   analysing a proteome of a tissue or a cell type. EMBRY proteins are
CC   useful as immunogens for preparing antibodies. Polynucleotides of the
CC   invention are useful for creating knockin humanised animals or transgenic
CC   animals to model human diseases. They are also used in gene therapy. The
CC   present sequence is human EMBRY-2 cDNA
XX
SQ   Sequence 2074 BP; 583 A; 567 C; 392 G; 531 T; 0 U; 1 Other;

  Query Match             86.0%;  Score 546.4;  DB 6;  Length 2074;
  Best Local Similarity   93.8%;  Pred. No. 5.9e-159;
  Matches  579;  Conservative    0;  Mismatches   37;  Indels    1;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||  || ||||||| |||
Db       1459 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1518

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||||||| ||| |||  ||||||
Db       1519 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAAATCTACCACGGACCC 1578

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||| |||||||| |||||| ||
Db       1579 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACTCCTCCCAAG 1638

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       1639 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 1698

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1699 CGTCAGTCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 1758

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||| |  ||||||||||||||||||||||||||||||| 
Db       1759 AGGACTAGCTGGATTTCCTAGGCCGATTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 1818

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              || ||||||| |||||| ||||||||||||||||||||||||| ||||||||||||||||
Db       1819 GCGTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAACCAATCAGAGAGCTC 1878

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       1879 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 1938

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
```

```
              ||||||||| ||||||||||||||| |  ||||||||||||||||||||||| || |  |
Db       1939 AGAGCACAGTGGGAGGGACAAGGATTGCAATATAAACCCAGGCATTCGAGCCAGC-ANAG 1997

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||| |||||||||||| |||| ||||||||| ||||||||||||||||||||||||||
Db       1998 CAACCGCCTTTGGGTCCCTTCCCCTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 2057

Qy        601 ATTAAATCTTGCAACTG 617
              |||||||||||||||||
Db       2058 ATTAAATCTTGCAACTG 2074


RESULT 7
AAV43219
ID   AAV43219 standard; cDNA; 1329 BP.
XX
AC   AAV43219;
XX
DT   29-DEC-1998  (first entry)
XX
DE   Multiple sclerosis associated retrovirus fragment 6.
XX
KW   Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW   env gene; rheumatoid arthritis-associated virus; ss.
XX
OS   Multiple sclerosis associated retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             2. .490
FT                   /*tag=  a
FT                   /product= "Encodes protein AAW71069"
FT                   /transl_except= (pos:77-79, appears to code for a stop
FT                   codon)
FT                   /transl_except= (pos:125-127, appears to code for a stop
FT                   codon)
FT                   /transl_except= (pos:137-139, appears to code for a stop
FT                   codon)
XX
PN   WO9823755-A1.
XX
PD   04-JUN-1998.
XX
PF   26-NOV-1997;   97WO-IB001482.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B;
XX
DR   WPI; 1998-322732/28.
DR   P-PSDB; AAW71069.
XX
PT   New nucleic acid from retroviruses - useful for diagnosis, prevention and
PT   treatment of, e.g. multiple sclerosis.
XX
PS   Disclosure; Page 187-188; 286pp; English.
XX
CC   The present sequence represents a multiple sclerosis (MS) associated
CC   retrovirus (MSRV) genomic fragment used in the method of the invention.
CC   The invention provides complete or partial genomic sequences of the MSRV-
CC   1 pol gene, gag gene and env gene, and polypeptides encoded by these
CC   genes. The invention also provides antibodies raised against the
CC   polypeptides. The genomic sequences, polypeptides and antibodies are also
CC   claimed useful for diagnosing infection by MS and rheumatoid arthritis-
CC   associated viruses, and also for prevention and treatment of infection
CC   with these viruses
XX
SQ   Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;

  Query Match             85.8%;  Score 545;  DB 2;  Length 1329;
  Best Local Similarity   92.4%;  Pred. No. 1.3e-158;
  Matches  572;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||| ||| ||||||||||||||||||||||||||||  || ||||||| |||
Db        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||| | |||||| ||||||||||||||||||| |||||||||||||
Db        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||| ||||||| |||||||||||||
Db        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||| ||||||||| ||||||||||| |||||||||||||||||||||||
Db        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||||||| |||||||||| ||||||||||| ||||||||| |||| ||||| |
Db        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856

Qy        601 ATTAAATCTTGCAACTGAA 619
              |||||||| |||||||| |
Db        857 ATTAAATCATGCAACTGCA 875


RESULT 8
AAX29704
ID   AAX29704 standard; DNA; 1329 BP.
XX
AC   AAX29704;
XX
DT   17-OCT-2003  (revised)
DT   27-AUG-2003  (revised)
DT   08-JUN-1999  (first entry)
XX
DE   Clone 5M6 from MSRV-1.
XX
KW   Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KW   rheumatoid polyarthritis; ss.
XX
OS   Viruses.
XX
PN   FR2765588-A1.
XX
PD   08-JAN-1999.
XX
PF   07-JUL-1997;   97FR-00008816.
XX
PR   07-JUL-1997;   97FR-00008816.
XX
PA   (INMR ) BIO MERIEUX.
```

```
XX
DR   WPI; 1999-098275/09.
DR   P-PSDB; AAW99554.
XX
PT   Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT   multiple sclerosis or rheumatoid polyarthritis.
XX
PS   Claim 1; Page 39-40; 83pp; French.
XX
CC   This sequence represents clone 5M6 from a novel multiple sclerosis
CC   related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC   prophylactic or therapeutic compositions to inhibit expression of a
CC   multiple sclerosis related virus and/or virus associated with rheumatoid
CC   polyarthritis. (Updated on 27-AUG-2003 to correct OS field.) (Updated on
CC   17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;

  Query Match             85.8%;  Score 545;  DB 2;  Length 1329;
  Best Local Similarity   92.4%;  Pred. No. 1.3e-158;
  Matches  572;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||| ||| ||||||||||||||||||||||||||||  || ||||||| |||
Db        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||| | |||||| ||||||||||||||||||| |||||||||||||
Db        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||| ||||||| |||||||||||||
Db        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||| ||||||||| ||||||||||| |||||||||||||||||||||||
Db        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||||||| |||||||||| ||||||||||| ||||||||| |||| ||||| |
Db        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856

Qy        601 ATTAAATCTTGCAACTGAA 619
              |||||||| |||||||| |
Db        857 ATTAAATCATGCAACTGCA 875


RESULT 9
ADG14849
ID   ADG14849 standard; cDNA; 1329 BP.
XX
AC   ADG14849;
XX
DT   26-FEB-2004  (first entry)
XX
DE   MSRV associated cDNA #11.
XX
KW   ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX
OS   Multiple sclerosis associated retrovirus.
XX
PN   US2003198647-A1.
XX
PD   23-OCT-2003.
XX
PF   03-APR-2002; 2002US-00114104.
XX
PR   26-NOV-1996;   96US-00756429.
PR   26-NOV-1997;   97US-00979847.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2004-032461/03.
XX
PT   New isolated nucleic acid and their fragments having the pol gene of a
PT   retrovirus, useful for diagnosing, preventing and/or treating multiple
PT   sclerosis and/or rheumatoid arthritis.
XX
PS   Disclosure; SEQ ID NO 108; 193pp; English.
XX
CC   The invention relates to an isolated nucleic acid which comprises the pol
CC   gene of a retrovirus associated with multiple sclerosis or rheumatoid
CC   arthritis. The methods and compositions of the present invention are
CC   useful for diagnosing, preventing and/or treating multiple sclerosis
CC   and/or rheumatoid arthritis. The present sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;

  Query Match             85.8%;  Score 545;  DB 12;  Length 1329;
  Best Local Similarity   92.4%;  Pred. No. 1.3e-158;
  Matches  572;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||| ||| ||||||||||||||||||||||||||||  || ||||||| |||
Db        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||| | |||||| ||||||||||||||||||| |||||||||||||
Db        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||| ||||||| |||||||||||||
Db        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||| ||||||||| ||||||||||| |||||||||||||||||||||||
Db          677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736

Qy          481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
                |||||||||||||| |||||||||| ||||||||||| ||||||||| |||| ||||| |
Db          737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796

Qy          541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
                ||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db          797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856

Qy          601 ATTAAATCTTGCAACTGAA 619
                |||||||| |||||||| |
Db          857 ATTAAATCATGCAACTGCA 875


RESULT 10
ACC46747
ID   ACC46747 standard; cDNA; 2046 BP.
XX
AC   ACC46747;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human dithp growth/development-associated protein-encoding cDNA.
XX
KW   Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW   cancer; cell proliferative disorder; autoimmune disorder;
KW   inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW   neurological disorder; gastrointestinal disorder; transport disorder;
KW   connective tissue disorder; drug screening; proteome analysis;
KW   gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW   disease model; toxicological testing; transcript imaging; growth;
KW   development; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200297031-A2.
XX
PD   05-DEC-2002.
XX
PF   27-MAR-2002; 2002WO-US010056.
XX
PR   28-MAR-2001; 2001US-0279619P.
PR   29-MAR-2001; 2001US-0280067P.
PR   29-MAR-2001; 2001US-0280068P.
PR   16-MAY-2001; 2001US-0291280P.
PR   17-MAY-2001; 2001US-0291829P.
PR   17-MAY-2001; 2001US-0291849P.
PR   19-JUN-2001; 2001US-0299428P.
PR   20-JUN-2001; 2001US-0299776P.
PR   20-JUN-2001; 2001US-0300001P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Daffo A,  Jones AL,  Tran AB,  Dahl CR,  Gietzen D,  Chinn J;
PI   Dufour GE,  Hillman JL,  Yu JY,  Tuason O,  Yap PE,  Amshey SR;
PI   Daughtery SC,  Dam TC,  Liu TF,  Nguyen DA,  Kleefeld Y,  Gerstin EH;
PI   Peralta CH,  David MH,  Lewis SA,  Chen AJ,  Panzer SR,  Harris B;
PI   Flores V,  Marwaha R,  Lo A,  Lan RY,  Urashka ME;
XX
DR   WPI; 2003-129518/12.
DR   P-PSDB; ABR41810.
XX
PT   Novel human diagnostic and therapeutic polypeptide useful for identifying
PT   test compound which specifically binds to a polypeptide encoded by human
PT   diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
PS   Claim 2; SEQ ID NO 668; 591pp; English.
XX
CC   The invention relates to novel human diagnostic and therapeutic
CC   polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC   proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC   polynucleotide sequences at least 90% identical to the dithp cDNA
CC   sequences of the invention; recombinant vectors, host cells and
CC   transgenic organisms comprising a dithp nucleic acid sequence; the
CC   recombinant production of DITHP proteins; antibodies specific for DITHP
CC   proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC   detecting dithp nucleotide and protein sequences; methods of screening
CC   for compounds which specifically bind a DITHP protein; and methods of
CC   assessing the toxicity of test compounds using a dithp hybridisation
CC   probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC   diagnosis of a wide variety of conditions including cancer and other cell
CC   proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC   viral, fungal or parasitic infections; hormonal disorders; metabolic
CC   disorders; neurological disorders; gastrointestinal disorders; transport
CC   disorders; and connective tissue disorders. They may also be used to
CC   screen for modulators of protein activity or gene expression. DITHP
CC   proteins can additionally be used in analysis of the proteome of a tissue
CC   or cell type and to induce antibodies. The dithp nucleic acids are
CC   additionally useful in somatic or germline gene therapy of the disorders
CC   mentioned above, as a source of antisense sequences, as a source of
CC   probes and primers, in genotyping and identification of individuals, in
CC   the generation of transgenic animal models of human disease or knock in
CC   humanised animals, in toxicological testing, and in transcript imaging.
CC   The present sequence represents a dithp cDNA encoding a DITHP protein
CC   which is associated with growth and development. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2046 BP; 587 A; 494 C; 391 G; 566 T; 0 U; 8 Other;

  Query Match             85.1%;  Score 540.2;  DB 8;  Length 2046;
  Best Local Similarity   94.4%;  Pred. No. 5e-157;
  Matches  560;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy            1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
                | ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db          571 CTCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAACTGTAAAACTA 630

Qy           61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                ||||| |||||||||||||||| ||||||| |||||||||||||||||||||   |||||
Db          631 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACCACAGACCC 690

Qy          121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
                ||||||| ||||||||||||||||||||||||||||||||| ||||||||||| ||||||
Db          691 CTGGACCAGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCCCCCGAG 750

Qy          181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
                |||||||||||||  |||||||||||||||||||||||||| ||||||||||| ||||||
Db          751 GAAATCTCAACTGTGCAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTACAGCGGT 810

Qy          241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                ||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db          811 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGAGGGCACTGAGAGAC 870

Qy          301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
                | |||||||||||||||||||||| || |||||||||||||| |||||||||||||||||
Db          871 AAGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCGTAGCTGGGAAGGTGACT 930

Qy          361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                |||||||||| |||||||||||||||| ||||||||||||||| ||||||||||||||||
Db          931 GCATCCACCTTTAAACATGGGGCTTGCCACTTAGCTCACACCCAACCAATCAGAGAGCTC 990

Qy          421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
                |||||||||||||||||||||||| |||||||||||||||||||||||| || |||||||
Db          991 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCACCTTTTGCCTG 1050

Qy          481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
                |||||||||| |||||||||||||||||||||||||||| ||||||| ||||||||| ||
Db         1051 AGAGCACAGCAGGAGGGACAAGGATCGGGATATAAACCCGGGCATTCAAGCCGGCAATGG 1110

Qy          541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATT 593
                ||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db         1111 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATT 1163
```

```
RESULT 11
ADB84403
ID   ADB84403 standard; DNA; 1329 BP.
XX
AC   ADB84403;
XX
DT   04-DEC-2003  (first entry)
XX
DE   MSRV-1 associated DNA sequence #21.
XX
KW   ds; multiple sclerosis; rheumatoid arthritis; gag; pol;
KW   reverse transcriptase; ribonuclease H.
XX
OS   Unidentified;.
XX
PN   US2003039664-A1.
XX
PD   27-FEB-2003.
XX
PF   26-NOV-1997;   97US-00979847.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (PERR/) PERRON H.
PA   (BESE/) BESEME F.
PA   (BEDI/) BEDIN F.
PA   (PARA/) PARANHOS-BACCALA G.
PA   (KOMU/) KOMURIAN-PRADEL F.
PA   (JOLI/) JOLIVET-REYNAUD C.
PA   (MAND/) MANDRAND B.
PA   (GARS/) GARSON J A.
PA   (TUKE/) TUKE P W.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2003-512253/48.
XX
PT   New isolated or purified nucleic acid associated with multiple sclerosis
PT   and/or rheumatoid arthritis, useful for detecting a virus associated with
PT   multiple sclerosis or rheumatoid arthritis in a biological sample.
XX
PS   Claim 31; Page 80; 193pp; English.
XX
CC   The invention relates to an isolated or purified nucleic acid from a
CC   virus associated with multiple sclerosis and/or rheumatoid arthritis,
CC   multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
CC   pol, gag or reverse transcriptase genes (or their fragments) encoding the
CC   proteins or defined peptides (including immunodominant peptides,
CC   antigenic peptides or conserved motifs).Also included are a process for
CC   detecting a virus associated with multiple sclerosis or rheumatoid
CC   arthritis in a biological sample, a nucleic acid probe for the detection
CC   of a virus associated with multiple sclerosis or rheumatoid arthritis, a
CC   primer for the amplification by polymerisation of a nucleic acid of a
CC   viral material associated with multiple sclerosis or rheumatoid
CC   arthritis, a polypeptide exhibiting an inhibitory activity on the
CC   proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC   and an antibody directed against the MSRV-1 virus obtained by
CC   immunologically reacting a human or animal body or cells with an
CC   immunogenic agent consisting of the antigenic polypeptide defined above.
CC   The nucleic acids are useful for detecting a biological sample a virus
CC   associated with multiple sclerosis or rheumatoid arthritis, or for
CC   detecting in a biological sample, the presence of or exposure to a virus
CC   associated with multiple sclerosis or rheumatoid arthritis. The present
CC   sequence is a claimed MSRV-associated sequence whose identity cannot be
CC   accurately determined. Note: The SEQ ID numbers for the sequences as
CC   displayed in the main body of the patent do not match the SEQ ID numbers
CC   in the sequence listing. Consequently those sequences mentioned in the
CC   claims may not be the sequences the authors intended to claim.
XX
SQ   Sequence 1329 BP; 339 A; 346 C; 282 G; 332 T; 0 U; 30 Other;
```

```
  Query Match             84.3%;  Score 535;  DB 9;  Length 1329;
  Best Local Similarity   90.8%;  Pred. No. 1.7e-155;
  Matches  562;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||| ||| |||||||||||||||||||||||||||   || ||||||| |||
Db        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAANTGAAGCTGTAAAGCTA 316

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||| | ||||||| ||||||||||||  ||| ||||||||||||
Db        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTANAATCTACCGTGGACCC 376

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||| | |||||| |||||||||||||||| || |||||||||||||
Db        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGNAGTCACCCCTCCCGAG 436

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||  | |||||||||  | |||||||||||||| ||
Db        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGNAAGCAGTTAGAGCAGT 496

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |||||||||||||||||||| |||||||||||||||||||||| ||| ||||||||||||
Db        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGNGGGTGGACTGAGAGAC 556

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||| ||||| | ||||||||||||| 
Db        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCNANCTGGGAAGGTGACC 616

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||| ||  ||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCNACCAATCAGAGAGCTC 676

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||| ||||||||| |||||||||||| |||||| ||||||||||||||||
Db        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCNAATCATCTATTGCCTG 736

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||||||| |||||||||| ||||||||||| |||||| ||| |||| ||||| |
Db        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCNTTCAAGCCAGCAACAG 796

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||| ||||||||| ||||||||| ||||||||||||||||
Db        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTNCACTCTATTTCACTCT 856

Qy        601 ATTAAATCTTGCAACTGAA 619
              |||||||| |||||||| |
Db        857 ATTAAATCATGCAACTGCA 875


RESULT 12
AAX77526
ID   AAX77526 standard; cDNA; 2946 BP.
XX
AC   AAX77526;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human secreted protein AJ172_2 cDNA.
XX
KW   Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW   bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW   cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW   haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;
KW   haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW   cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
OS   Homo sapiens.
XX
PN   WO9926972-A1.
XX
PD   03-JUN-1999.
XX
PF   17-NOV-1998;   98WO-US024614.
```

```
XX
PR   21-NOV-1997;   97US-00976110.
PR   18-MAY-1998;   98US-00080478.
PR   20-OCT-1998;   98US-00175928.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M;
XX
DR   WPI; 1999-357813/30.
DR   P-PSDB; AAY08622.
XX
PT   New polynucleotides encoding secreted proteins.
XX
PS   Claim 13a; Page 100-101; 142pp; English.
XX
CC   This invention describes novel human secreted proteins encoded by
CC   polynucleotides isolated from human adult testes, adult brain, adult
CC   blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC   libraries. The products of the invention are predicted to have biological
CC   activities which would make them suitable for treating, preventing or
CC   ameliorating medical conditions in humans and animals, although no
CC   supporting data is given. Suggested activities include nutritional
CC   activity, cytokine and cell proliferation/differentiation activity,
CC   immune stimulating (e.g. as vaccines) or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, and tumour
CC   inhibition activity. The polynucleotides are also stated to be useful for
CC   gene therapy
XX
SQ   Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

  Query Match             84.1%;  Score 533.8;  DB 2;  Length 2946;
  Best Local Similarity   91.5%;  Pred. No. 5.9e-155;
  Matches  581;  Conservative    0;  Mismatches   42;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2323 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2382

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2383 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2430

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2431 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2490

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2491 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2550

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2551 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGACTGAGAGAC 2610

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       2611 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2670

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2671 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2730

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       2731 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2790

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2791 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2850

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2851 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2910

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       2911 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2945


RESULT 13
AAZ59468
ID   AAZ59468 standard; cDNA; 2946 BP.
XX
AC   AAZ59468;
XX
DT   11-APR-2000  (first entry)
XX
DE   Human secreted protein AJ172_2 polynucleotide sequence.
XX
KW   Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW   placental pathology; metastasis inhibition; nutritional activity;
KW   immune stimulator; haematopoiesis regulator; tissue growth;
KW   tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9960020-A1.
XX
PD   25-NOV-1999.
XX
PF   17-MAY-1999;   99WO-US010915.
XX
PR   18-MAY-1998;   98US-00080478.
PR   20-OCT-1998;   98US-00175928.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Mi S,  Treacy M;
XX
DR   WPI; 2000-116311/10.
DR   P-PSDB; AAY67313.
XX
PT   New polynucleotides encoding secreted cDNA libraries, used to develop
PT   products for the diagnosis and treatment of neoplastic disease.
XX
PS   Claim 14; Page 107-108; 149pp; English.
XX
CC   This is the human secreted protein AJ172_2 nucleotide sequence, obtained
CC   from a human adult testes cDNA library. The invention relates to secreted
CC   human and murine proteins. The polynucleotides and proteins are predicted
CC   to have biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals. Detection of the levels of the proteins can be used for the
CC   diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
CC   which modulate the expression or function of the proteins may be used for
CC   treating a neoplastic disease and inhibiting metastasis. Other suggested
CC   activities include nutritional activity (e.g. in feeds), cytokine and
CC   cell proliferation/differentiation activity, immune stimulating (e.g. as
CC   vaccines) or suppressing activity, haematopoiesis regulating activity,
CC   tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC   invasion suppressor activity, and tumour inhibition activity. The
CC   polynucleotide sequences are also stated to be useful for gene therapy
XX
SQ   Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

  Query Match             84.1%;  Score 533.8;  DB 3;  Length 2946;
  Best Local Similarity   91.5%;  Pred. No. 5.9e-155;
  Matches  581;  Conservative    0;  Mismatches   42;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2323 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2382

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2383 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2430

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2431 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2490

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2491 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2550

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2551 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2610

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       2611 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2670

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2671 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2730

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       2731 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2790

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2791 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2850

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2851 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2910

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       2911 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2945


RESULT 14
ADC38776
ID   ADC38776 standard; cDNA; 2946 BP.
XX
AC   ADC38776;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human cDNA encoding a secreted protein #63.
XX
KW   ss; gene; immune disorder; severe combined immunodeficiency; SCID;
KW   autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW   rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;
KW   lymphoid cell deficiency; osteoporosis; osteoarthritis;
KW   peripheral nervous system disease; peripheral neuropathy;
KW   Alzheimer's disease; Parkinson's disease; coagulation disorder;
KW   inflammatory disease; systemic inflammatory response syndrome; SIRS;
KW   ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;
KW   hypersensitivity; regeneration; neural cell proliferation; fertility;
KW   tumour; chemokine; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   US2002193567-A1.
XX
PD   19-DEC-2002.
```

```
XX
PF   02-APR-2002; 2002US-00114893.
XX
PR   11-AUG-1995;   95US-00514014.
PR   05-APR-1996;   96US-00628364.
PR   19-APR-1996;   96US-00635311.
PR   07-JUN-1996;   96US-00659224.
PR   17-JUN-1996;   96US-00664596.
PR   09-JUL-1996;   96US-00677231.
PR   26-JUL-1996;   96US-00686878.
PR   23-AUG-1996;   96US-00701819.
PR   27-SEP-1996;   96US-00721488.
PR   27-SEP-1996;   96US-00721798.
PR   27-SEP-1996;   96US-00721923.
PR   27-SEP-1996;   96US-00721926.
PR   25-OCT-1996;   96US-00738367.
PR   30-OCT-1996;   96US-00739775.
PR   13-JAN-1997;   97US-00783395.
PR   10-APR-1997;   97US-00833823.
PR   02-JUN-1997;   97US-00867677.
PR   05-SEP-1997;   97US-00924838.
PR   06-OCT-1999;   99US-00413232.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR,  Spaulding V,  Carlin-Duckett M;
PI   Kelleher K;
XX
DR   WPI; 2003-657236/62.
DR   P-PSDB; ADC38777.
XX
PT   Proteins AZ3021 encoded by clone AZ3021 from human adult colon, and
PT   BD12716 encoded by clone BD12716 from human fetal kidney cDNA library,
PT   useful for treating e.g. multiple sclerosis and rheumatoid arthritis.
XX
PS   Disclosure; SEQ ID NO 134; 412pp; English.
XX
CC   The invention relates to a protein comprising fully defined AZ302 1
CC   protein or BD127 1 6 protein. The polynucleotides are useful for
CC   expressing recombinant proteins for analysis and are also useful as
CC   chromosome markers or tags to identify chromosomes or to map related gene
CC   positions. The proteins are useful as amino acid supplement, carbon
CC   source, nitrogen source and carbohydrate source. The proteins are useful
CC   for treating various immune deficiencies and disorders (e.g. severe
CC   combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC   sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic
CC   reactions (e.g. asthma), myeloid or lymphoid cell deficiencies,
CC   osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g.
CC   peripheral neuropathy, Alzheimer's disease, Parkinson's disease),
CC   coagulation disorders, inflammatory diseases (e.g. systemic inflammatory
CC   response syndrome (SIRS), ischaemia-reperfusion injury, Crohn's disease),
CC   anaphylaxis and hypersensitivity. Proteins are also useful for inducing
CC   tumour immunity, for inducing bone, cartilage, tendon, ligament and/or
CC   nerve growth or regeneration, for proliferating neural cells and for
CC   regenerating nerve and brain tissue, for inducing fertility and for
CC   inhibiting tumour growth. Proteins are also useful as chemokine for
CC   mammalian cells (e.g., monocytes, fibroblasts, neutrophils), and also
CC   useful as inhibitors of receptor/ligand interactions. The present
CC   sequence represents cDNA encoding a human secreted protein.
XX
SQ   Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

  Query Match             84.1%;  Score 533.8;  DB 10;  Length 2946;
  Best Local Similarity   91.5%;  Pred. No. 5.9e-155;
  Matches  581;  Conservative    0;  Mismatches   42;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2323 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2382

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2383 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2430

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2431 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2490

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2491 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2550

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2551 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2610

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       2611 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2670

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2671 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2730

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       2731 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2790

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2791 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2850

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2851 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2910

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       2911 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2945


RESULT 15
ABK89296/c
ID   ABK89296 standard; DNA; 326014 BP.
XX
AC   ABK89296;
XX
DT   16-OCT-2002  (first entry)
XX
DE   Human gene for novel serine/threonine serine kinase.
XX
KW   Human; ds; gene; serine/threonine protein kinase; inflammation; cancer;
KW   arteriosclerosis; psoriasis; SNF kinase; transgenic; chromosome 3; SNP;
KW   single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(1139. .1140,GGT)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1399,G)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1871. .1873,TG)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1875. .1877,TA)
FT                   /*tag=  al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             3000. .323016
FT                   /*tag=  a
FT                   /product= "Serine/threonine protein kinase"
FT   exon            3000. .3169
FT                   /*tag=  b
FT                   /number= 1
FT   intron          3170. .53737
FT                   /*tag=  c
FT                   /number= 1
FT   variation       replace(3180,G)
FT                   /*tag=  am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7590,A)
FT                   /*tag=  an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9315,C)
FT                   /*tag=  ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9572,A)
FT                   /*tag=  ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11004,G)
FT                   /*tag=  aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11379. .11380,AAA)
FT                   /*tag=  ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12478,A)
FT                   /*tag=  as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12486. .12487,TTT)
FT                   /*tag=  at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13904,G)
FT                   /*tag=  au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25444. .25446,TT)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26805,A)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28010,R)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28107,Y)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28178,R)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28356,T)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29535,G)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31808,T)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32807,C)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(34531,G)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35318,G)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35463,C)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35477,T)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37050,G)
FT                   /*tag=  bi
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37470,G)
FT                   /*tag=  bj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37481,A)
FT                   /*tag=  bk
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37499,T)
FT                   /*tag=  bl
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(40178. .40179,GRA)
FT                   /*tag=  bm
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(40787,A)
FT                   /*tag=  bn
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(44582,G)
FT                   /*tag=  bo
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(45615,A)
FT                   /*tag=  bp
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(46091,T)
FT                   /*tag=  bq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(47844,C)
FT                   /*tag=  br
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(48013,T)
FT                   /*tag=  bs
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(48072,A)
FT                   /*tag=  bt
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(50802,T)
FT                   /*tag=  bu
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(53305,A)
FT                   /*tag=  bv
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(53378. .53380,GG)
FT                   /*tag=  bw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            53738. .53903
FT                   /*tag=  d
FT                   /number= 2
FT   intron          53904. .83114
FT                   /*tag=  e
FT                   /number= 2
FT   variation       replace(56233,G)
FT                   /*tag=  bx
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(56453,C)
FT                   /*tag=  by
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(59300,T)
FT                   /*tag=  bz
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(60998,A)
FT                   /*tag=  ca
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63013,A)
FT                   /*tag=  cb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63293,A)
FT                   /*tag=  cc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64181,T)
FT                   /*tag=  cd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64191,A)
FT                   /*tag=  ce
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(68992,C)
```

```
FT                   /*tag=  cf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(76246,Y)
FT                   /*tag=  cg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(77368,G)
FT                   /*tag=  ch
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(78043,A)
FT                   /*tag=  ci
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(79395,T)
FT                   /*tag=  cj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(79474,G)
FT                   /*tag=  ck
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(79992,T)
FT                   /*tag=  cl
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(81065,C)
FT                   /*tag=  cm

  Query Match             83.9%;  Score 532.8;  DB 6;  Length 326014;
  Best Local Similarity   90.5%;  Pred. No. 1.2e-153;
  Matches  583;  Conservative    0;  Mismatches   52;  Indels    9;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||| |||  || || ||||||||
Db     170822 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAAAATTGAAGCTCTAAAACTA 170763

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||||||||||||||||||| ||||| | ||||||||||||||||| ||||  |||||
Db     170762 CAAATTGTTCTTCAAATGGAGCCCCAGACGCAGTCCATGACTAAGATCTACCGCAGACCC 170703

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db     170702 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCTCCAGAG 170643

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||| |||||| |||| ||||||||||||| ||||||||||||||| ||
Db     170642 GAAATCTCAACTGCATAACCCCCACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 170583

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              |||| |||| |||||||||||||||||||||||||||||||||||||| ||| ||||||
Db     170582 CATCGGCCATCCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGTACTCAGAGAC 170523

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||||
Db     170522 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 170463

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
              || || |||| |||||  |||||||||||||||||||||||| ||||||||
Db     170462 GCTTCTACCTTTAAACCCGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGGTAGGAA 170403

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              |||||||||||||||||| ||| ||||| |||| ||||||||||||||||||||||||||
Db     170402 AGAGAGCTCACTAAAATGGTAACTAGGCTAAAACAGGAGGTAAAGAAATAGCCAATCATC 170343

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              ||| |||||||| ||||| ||||||||||| ||| ||| |||||||||||||||| ||||
Db     170342 TATCGCCTGAGAACACAGTGGGAGGGACAATGATTGGGTTATAAACCCAGGCATTGGAGC 170283

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              | |||| ||| |||    |||||||||||||||||||||||| |||||||||||||||||
Db     170282 CAGCAATGGCTACCATTTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTA 170223

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||| | ||||||||||||||| |
Db     170222 TTTCACTCTATTAAATCTTGCAACAGCAAAAAAAAAAAAAAAGA 170179


RESULT 16
```

```
ADQ94981/c
ID   ADQ94981 standard; DNA; 326014 BP.
XX
AC   ADQ94981;
XX
DT   23-SEP-2004  (first entry)
XX
DE   Human kinase genomic DNA.
XX
KW   Kinase; diagnosis; cancer; tissue growth abnormality;
KW   drug screening assay; cytostatic; gene therapy; human; SNP;
KW   single nucleotide polymorphism; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(1138..1140,GT)
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1399,G)
FT                   /*tag= b
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1871..1873,TG)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1875..1877,TA)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   CDS             3000..323016
FT                   /*tag= f
FT                   /product= "Human kinase protein"
FT                   /transl_except= (pos:3024..3026, aa:Arg)
FT                   /transl_except= (pos:142261..142263, aa:Glu)
FT                   /transl_except= (pos:142261..144102, aa:Glu-Ser)
FT   exon            3000..3169
FT                   /*tag= e
FT   intron          3170..53737
FT                   /*tag= g
FT   variation       replace(3180,A)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(7590,A)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9315,C)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9572,A)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(11004,G)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(11378..11380,AA)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12478,A)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12485..12487,TC)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13904,G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(25444..25446,TT)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(26805,A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28010,R)
FT                   /*tag= s
```

```
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28107,Y)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28178,R)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28356,T)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(29535,G)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(31808,T)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(32807,C)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(34531,G)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35318,G)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35463,G)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35477,T)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37050,G)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37470,G)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37481,A)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37499,T)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(40177..40179,AA)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(40178,A)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(40787,A)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(44582,G)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(45615,A)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46091,T)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(47844,C)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(48013,T)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(48072,A)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(50802,T)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(53305,A)
FT                   /*tag=  ar
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(53378. .53380,GG)
FT                   /*tag=  as
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            53738. .53903
FT                   /*tag=  at
FT   intron          53904. .83114
FT                   /*tag=  au
FT   variation       replace(56233,G)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(56453,C)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(59300,T)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(60998,A)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(63013,A)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(63293,A)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(64181,T)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(64191,A)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(68992,C)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(76246,Y)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(77368,G)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(78043,A)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(79395,T)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(79474,G)
FT                   /*tag=  bi
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(79992,T)
FT                   /*tag=  bj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"

  Query Match             83.9%;  Score 532.8;  DB 12;  Length 326014;
  Best Local Similarity   90.5%;  Pred. No. 1.2e-153;
  Matches  583;  Conservative    0;  Mismatches   52;  Indels    9;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||| |||  || || ||||||||
Db     170822 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAAAATTGAAGCTCTAAAACTA 170763

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||||||||||||||||||| ||||| | ||||||||||||||||| ||||  |||||
Db     170762 CAAATTGTTCTTCAAATGGAGCCCCAGACGCAGTCCATGACTAAGATCTACCGCAGACCC 170703

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db     170702 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCTCCAGAG 170643

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
```

```
              ||||||||||||||| |||||| |||| ||||||||||||| ||||||||||||||| ||
Db     170642 GAAATCTCAACTGCATAACCCCCACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 170583

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              |||| |||| |||||||||||||||||||||||||||||||||||||| ||| ||||||
Db     170582 CATCGGCCATCCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGTACTCAGAGAC 170523

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||||
Db     170522 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 170463

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
              || || |||| |||||  |||||||||||||||||||||||| ||||||||
Db     170462 GCTTCTACCTTTAAACCCGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGGTAGGAA 170403

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              |||||||||||||||||| ||| ||||| |||| ||||||||||||||||||||||||||
Db     170402 AGAGAGCTCACTAAAATGGTAACTAGGCTAAAACAGGAGGTAAAGAAATAGCCAATCATC 170343

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              ||| |||||||| ||||| ||||||||||| ||| ||| |||||||||||||||| ||||
Db     170342 TATCGCCTGAGAACACAGTGGGAGGGACAATGATTGGGTTATAAACCCAGGCATTGGAGC 170283

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              | |||| ||| |||   ||||||||||||||||||||||||| |||||||||||||||||
Db     170282 CAGCAATGGCTACCATTTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTA 170223

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||| | ||||||||||||||| |
Db     170222 TTTCACTCTATTAAATCTTGCAACAGCAAAAAAAAAAAAAAAAGA 170179


RESULT 17
AAF55630
ID   AAF55630 standard; DNA; 2781 BP.
XX
AC   AAF55630;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
KW   Envelope protein; HERV; syncytia formation; placental development;
KW   syncytia; cancer; cell adhesion; ss.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             762. .2378
FT                   /*tag=  a
FT                   /product= "envelope protein"
XX
PN   WO200116171-A1.
XX
PD   08-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-FR002429.
XX
PR   01-SEP-1999;    99FR-00011141.
PR   15-SEP-1999;    99FR-00011793.
XX
PA   (INMR ) BIO MERIEUX.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI   Mallet F,  Cosset F,  Blond J,  Lavillette D,  Bouton O,  Ruggieri A;
XX
DR   WPI; 2001-226676/23.
DR   P-PSDB; AAB67652.
XX
PT   Detecting expression of human endogenous retrovirus envelope protein in
PT   cells of a tissue or culture, from its ability to induce syncytia.
XX
PS   Disclosure; Page 44-45; 57pp; French.
```

```
XX
CC   The present sequence encodes a human endogenous retrovirus envelope
CC   protein. The specification describes a method for detecting expression of
CC   an envelope protein from a human endogenous retrovirus (HERV), in cells,
CC   of a tissue or culture. The method comprises detecting syncytia formation
CC   due to the fusogenic properties of the envelope protein. Envelope
CC   polypeptides and polynucleotides are used to produce therapeutic or
CC   prophylactic compositions, particularly for treatment of cancer, to
CC   correct defects in placental development (or other natural formation of
CC   other types of syncytia), and to promote adhesion of cells in grafts or
CC   cellular repair processes. Expression of sequences antisense to the
CC   polynucleotide are used to prevent formation of syncytia
XX
SQ   Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 U; 0 Other;

  Query Match             83.8%;  Score 532.2;  DB 5;  Length 2781;
  Best Local Similarity   91.3%;  Pred. No. 1.8e-154;
  Matches  580;  Conservative    0;  Mismatches   43;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2157 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2216

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2217 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2264

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2265 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2324

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2325 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2384

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2385 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2444

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||| 
Db       2445 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2504

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2505 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2564

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db       2565 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2624

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2625 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2684

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2685 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2744

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       2745 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2779


RESULT 18
AAX25660
ID   AAX25660 standard; cDNA to mRNA; 1136 BP.
XX
AC   AAX25660;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Human endogenous retrovirus W clone cl.C4C5.
```

```
XX
KW   Clone; human endogenous retrovirus; genome; autoimmune disease;
KW   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW   disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9902696-A1.
XX
PD   21-JAN-1999.
XX
PF   06-JUL-1998;   98WO-FR001442.
XX
PR   07-JUL-1997;   97FR-00008815.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Beseme F,  Blond J,  Bouton O,  Mandrand B,  Mallet F;
XX
DR   WPI; 1999-120897/10.
XX
PT   New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT   exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT   disease, and abnormal or failed pregnancy.
XX
PS   Claim 1; Page 59-60; 106pp; French.
XX
CC   This sequence represents clone cl.C4C5 of the human endogenous retrovirus
CC   (HERV) W genome. The nucleic acids, their fragments or peptides encoded
CC   by them are markers of autoimmune disease (e.g. multiple sclerosis,
CC   rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC   dependent diabetes and related pathologies) and of abnormal or
CC   unsuccessful pregnancy and can be used as chromosomal markers for
CC   susceptibility to these conditions, or proximity markers of genes
CC   associated with this susceptibility
XX
SQ   Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 U; 0 Other;

  Query Match             82.8%;  Score 525.8;  DB 2;  Length 1136;
  Best Local Similarity   90.7%;  Pred. No. 1.1e-152;
  Matches  576;  Conservative    0;  Mismatches   47;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        510 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 569

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        570 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 617

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db        618 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 677

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        678 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 737

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db        738 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 797

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||| 
Db        798 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 857

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||| ||||||| |||||||||||||||||
Db        858 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGTTCACACCTGACCAATCAGAGAGCTC 917

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||| ||
Db        918 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCATG 977
```

```
Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db        978 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 1037

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       1038 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 1097

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||| |||  ||||||||||||||||
Db       1098 ATTAAATCTTGCAGCTGCGAAAAAAAAAAAAAAAA 1132


RESULT 19
AAA59210
ID   AAA59210 standard; DNA; 1136 BP.
XX
AC   AAA59210;
XX
DT   07-NOV-2000  (first entry)
XX
DE   3' pol gene and 3' non coding sequences of HERV-W from human genome.
XX
KW   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW   gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200043521-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-FR000144.
XX
PR   21-JAN-1999;   99FR-00000888.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Mallet F,  Voisset C;
XX
DR   WPI; 2000-499229/44.
XX
PT   New nucleic acid from human endogenous retrovirus, useful e.g. for
PT   diagnosis of autoimmune disease and complications of pregnancy, contains
PT   at least part of the gag gene.
XX
PS   Disclosure; Page 46; 53pp; French.
XX
CC   The present sequence represents an endogenetic retroviral nucleic acid
CC   fragment, which is associated with an autoimmune disease, and is
CC   integrated into the human genome. The fragment is originally derived from
CC   a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W
CC   retrovirus is associated with autoimmune disease, failure of pregnancy or
CC   disorders of pregnancy. The nucleic acid fragment, or proteins derived
CC   from it, are useful for diagnosis of autoimmune disease (specifically
CC   multiple sclerosis) and for monitoring pregnancy. The nucleic acid
CC   fragments may also be used for in situ labelling of isolated chromosomes,
CC   while the transcription product can be used to study or monitor T cell
CC   proliferation in vitro
XX
SQ   Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 U; 0 Other;

  Query Match             82.8%;  Score 525.8;  DB 3;  Length 1136;
  Best Local Similarity   90.7%;  Pred. No. 1.1e-152;
  Matches  576;  Conservative    0;  Mismatches   47;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        510 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 569

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        570 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 617

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db        618 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 677

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        678 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 737

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db        738 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGACTGAGAGAC 797

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db        798 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 857

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||| ||||||| |||||||||||||||||
Db        858 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGTTCACACCTGACCAATCAGAGAGCTC 917

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||| ||
Db        918 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCATG 977

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db        978 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 1037

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       1038 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 1097

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||| |||  ||||||||||||||||
Db       1098 ATTAAATCTTGCAGCTGCGAAAAAAAAAAAAAAAA 1132


RESULT 20
AAX25661
ID   AAX25661 standard; cDNA to mRNA; 2782 BP.
XX
AC   AAX25661;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Human endogenous retrovirus W clone cl.PH74.
XX
KW   Clone; human endogenous retrovirus; genome; autoimmune disease;
KW   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW   disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9902696-A1.
XX
PD   21-JAN-1999.
XX
PF   06-JUL-1998;   98WO-FR001442.
XX
PR   07-JUL-1997;   97FR-00008815.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Beseme F,  Blond J,  Bouton O,  Mandrand B,  Mallet F;
XX
DR   WPI; 1999-120897/10.
XX
PT   New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT   exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT   disease, and abnormal or failed pregnancy.
XX
```

```
PS   Claim 1; Page 60-63; 106pp; French.
XX
CC   This sequence represents clone cl.PH74 of the human endogenous retrovirus
CC   (HERV) W genome. The nucleic acids, their fragments or peptides encoded
CC   by them are markers of autoimmune disease (e.g. multiple sclerosis,
CC   rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC   dependent diabetes and related pathologies) and of abnormal or
CC   unsuccessful pregnancy and can be used as chromosomal markers for
CC   susceptibility to these conditions, or proximity markers of genes
CC   associated with this susceptibility
XX
SQ   Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;

  Query Match             82.3%;  Score 522.6;  DB 2;  Length 2782;
  Best Local Similarity   90.4%;  Pred. No. 1.7e-151;
  Matches  574;  Conservative    0;  Mismatches   49;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||| ||||||||||||| |||||||||||||||||| || |||| ||||||
Db       2158 CCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTGAAACTA 2217

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2218 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2265

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2266 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2325

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2326 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2385

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2386 GGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2445

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||| ||||||||| |||||||||||| 
Db       2446 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCTAGGTGGGAAGGTGACC 2505

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2506 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2565

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | ||||||||||||||||||||||||| |||||||||
Db       2566 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATTTATTGCCTG 2625

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2626 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTCGAGCCGGCAACGG 2685

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2686 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2745

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       2746 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2780


RESULT 21
AAA59211
ID   AAA59211 standard; DNA; 2782 BP.
XX
AC   AAA59211;
XX
DT   07-NOV-2000  (first entry)
XX
DE   5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.
XX
KW   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW   gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200043521-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-FR000144.
XX
PR   21-JAN-1999;   99FR-00000888.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Mallet F,  Voisset C;
XX
DR   WPI; 2000-499229/44.
XX
PT   New nucleic acid from human endogenous retrovirus, useful e.g. for
PT   diagnosis of autoimmune disease and complications of pregnancy, contains
PT   at least part of the gag gene.
XX
PS   Disclosure; Page 46-47; 53pp; French.
XX
CC   The present sequence represents an endogenetic retroviral nucleic acid
CC   fragment, which is associated with an autoimmune disease, and is
CC   integrated into the human genome. The fragment is originally derived from
CC   a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W
CC   retrovirus is associated with autoimmune disease, failure of pregnancy or
CC   disorders of pregnancy. The nucleic acid fragment, or proteins derived
CC   from it, are useful for diagnosis of autoimmune disease (specifically
CC   multiple sclerosis) and for monitoring pregnancy. The nucleic acid
CC   fragments may also be used for in situ labelling of isolated chromosomes,
CC   while the transcription product can be used to study or monitor T cell
CC   proliferation in vitro
XX
SQ   Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;

  Query Match             82.3%;  Score 522.6;  DB 3;  Length 2782;
  Best Local Similarity   90.4%;  Pred. No. 1.7e-151;
  Matches  574;  Conservative    0;  Mismatches   49;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||| ||||||||||||| |||||||||||||||||| || |||| ||||||
Db       2158 CCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTGAAACTA 2217

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2218 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2265

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2266 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2325

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2326 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2385

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2386 GGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2445

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||| ||||||||| |||||||||||| 
Db       2446 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCTAGGTGGGAAGGTGACC 2505

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2506 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2565

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | ||||||||||||||||||||||||| |||||||||
Db       2566 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATTTATTGCCTG 2625
```

```
Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2626 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTCGAGCCGGCAACGG 2685

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2686 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2745

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       2746 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2780


RESULT 22
AAH20069
ID   AAH20069 standard; DNA; 2782 BP.
XX
AC   AAH20069;
XX
DT   08-AUG-2001  (first entry)
XX
DE   HERV-W envelope protein G encoding nucleic acid.
XX
KW   Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW   envelope protein; multiple sclerosis-related superantigen; vaccine;
KW   surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW   antisense-therapy; autoimmune disorder; ds.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1. .762
FT                   /*tag=  a
FT   CDS             763. .2379
FT                   /*tag=  b
FT                   /product= "HERV-W envelope protein G"
FT                   /transl_except= (pos:790. .792,aa:Phe)
FT                   /transl_except= (pos:793. .795,aa:Thr)
FT                   /transl_except= (pos:812. .814,aa:Leu)
FT                   /transl_except= (pos:818. .820,aa:Ser)
FT                   /transl_except= (pos:862. .864,aa:Tyr)
FT                   /transl_except= (pos:865. .867,aa:Gln)
FT                   /transl_except= (pos:1174. .1176,aa:Arg)
FT                   /transl_except= (pos:1441. .1443,aa:Leu)
FT                   /transl_except= (pos:1903. .1905,aa:Gln)
FT                   /transl_except= (pos:2017. .2019,aa:Lys)
FT                   /transl_except= (pos:2026. .2028,aa:Arg)
FT                   /transl_except= (pos:2044. .2046,aa:Arg)
FT                   /transl_except= (pos:2089. .2091,aa:Gln)
FT                   /transl_except= (pos:2170. .2172,aa:Asn)
FT   3'UTR           2380. .2782
FT                   /*tag=  c
XX
PN   WO200131021-A1.
XX
PD   03-MAY-2001.
XX
PF   30-OCT-2000; 2000WO-EP010659.
XX
PR   28-OCT-1999;   99EP-00402690.
XX
PA   (UYGE-) UNIV GENEVE.
XX
PI   Conrad B,  Mach B;
XX
DR   WPI; 2001-316336/33.
DR   P-PSDB; AAB75138.
XX
PT   New human retrovirus HERV-W ENV proteins/peptides having superantigen
PT   activity useful for diagnosing and treating multiple sclerosis.
XX
PS   Claim 13; Fig 9; 94pp; English.
XX
CC   On the basis of the PBS t-RNA motif used for the classification of human
CC   endogenous retrovirus (HERVs) the full length endogenous provirus which
CC   was been located on the long arm of human chromosome 7 (7q21-22) has been
CC   designated HERV-W. The present invention describes proteins or peptides
CC   (I) having superantigen (SAg) activity comprising the ENV protein (ENV)
CC   of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC   have neuroprotective activity, and can be used in: vaccines; antisense-
CC   therapy; and HERV-W SAg activity-inhibitors. (I) and encoding DNA/RNA are
CC   useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC   disorders. (I) are also useful for identifying substances (and optionally
CC   recovering) capable of binding to a retroviral superantigen associated
CC   with MS, substances capable of blocking SAg activity and substances
CC   capable of blocking transcription or translation of HERV-W retroviral
CC   superantigen. A protein or peptide derived from (I), modified to be
CC   devoid of SAg activity and being capable of generating an immune response
CC   against HERV-W retroviral SAg is useful in therapy. Nucleic acid
CC   molecules encoding (I) are useful as vaccines against MS. Substances
CC   capable of blocking SAg activity, capable of binding to a retroviral
CC   superantigen associated with MS, or capable of blocking transcription or
CC   translation of HERV-W retroviral superantigen for use in treating or
CC   preventing MS, obtained using (I) are useful for the treatment and
CC   prevention of MS. (I) and nucleic acids encoding them are useful for
CC   diagnosing autoimmune disease. The present sequence encodes the
CC   specifically claimed envelope protein of HERV-W designated G
XX
SQ   Sequence 2782 BP; 741 A; 768 C; 564 G; 709 T; 0 U; 0 Other;

  Query Match             82.3%;  Score 522.6;  DB 5;  Length 2782;
  Best Local Similarity   90.4%;  Pred. No. 1.7e-151;
  Matches  574;  Conservative    0;  Mismatches   49;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||| ||||||||||||| |||||||||||||||||| || |||| ||||||
Db       2158 CCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTGAAACTA 2217

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2218 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2265

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2266 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2325

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2326 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2385

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2386 GGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2445

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||| ||||||||| ||||||||||||
Db       2446 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCTAGGTGGGAAGGTGACC 2505

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2506 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2565

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | ||||||||||||||||||||||||| |||||||||
Db       2566 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATTTATTGCCTG 2625

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2626 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTCGAGCCGGCAACGG 2685

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2686 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2745

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
```

```
              ||||||||||||||||| |||||||||||||||||
Db       2746 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2780

RESULT 23
AAD24195
ID   AAD24195 standard; cDNA; 2930 BP.
XX
AC   AAD24195;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Human syncytin cDNA.
XX
KW   Human; syncytin; preeclampsia; gestational trophoblast disorder;
KW   choriocarcinoma; hydatiform mole; placental site tumour; abortion;
KW   envelope gene; human endogenous defective retrovirus; HERV-W; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             930. .2546
FT                   /*tag=  a
FT                   /product= "Syncytin"
XX
PN   WO200204678-A2.
XX
PD   17-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-US021719.
XX
PR   07-JUL-2000; 2000US-0216657P.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Keith JC,  Mccoy JM,  Mi S;
XX
DR   WPI; 2002-171727/22.
DR   P-PSDB; AAE14540.
XX
PT   Identifying a compound for treating a subject with or at risk of
PT   developing preeclampsia, comprises determining whether the expression or
PT   activity of syncytin in the cell is modulated in the presence of a test
PT   compound.
XX
PS   Disclosure; Page 39-42; 43pp; English.
XX
CC   The invention relates to identifying compounds which are modulators of
CC   syncytin expression. The syncytin modulators are useful in diagnosis and
CC   treatment of preeclampsia and gestational trophoblast disorders (e.g.
CC   choriocarcinoma, hydatiform mole, placental site tumour and missed/
CC   incomplete abortion). Syncytin is a human gene derived from the envelope
CC   gene of human endogenous defective retrovirus, HERV-W. The present
CC   invention is based partly on the discovery that syncytin expression is
CC   dramatically reduced in preeclampsia, and is also mis-localised to the
CC   apical syncytiotrophoblast membrane. The present sequence is human
CC   syncytin cDNA
XX
SQ   Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 U; 0 Other;

  Query Match             81.5%;  Score 517.4;  DB 6;  Length 2930;
  Best Local Similarity   91.4%;  Pred. No. 7.4e-150;
  Matches  564;  Conservative    0;  Mismatches   41;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2325 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2384

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2385 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2432

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2433 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2492

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2493 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2552

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2553 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2612

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||| 
Db       2613 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2672

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2673 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2732

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       2733 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2792

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2793 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2852

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2853 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2912

Qy        601 ATTAAATCTTGCAACTG 617
              |||||||||||||||||
Db       2913 ATTAAATCTTGCAACTG 2929


RESULT 24
ABL61744
ID   ABL61744 standard; DNA; 56093 BP.
XX
AC   ABL61744;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Colon adenocarcinoma related gene sequence SEQ ID NO:81.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 81; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 U; 0 Other;

  Query Match             81.3%;  Score 516.2;  DB 6;  Length 56093;
  Best Local Similarity   91.1%;  Pred. No. 7.4e-149;
  Matches  564;  Conservative    0;  Mismatches   43;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db      37274 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 37333

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db      37334 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 37381

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db      37382 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 37441

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db      37442 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 37501

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| ||||||||||||||
Db      37502 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGACTGAGAGAC 37561

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db      37562 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 37621

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db      37622 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 37681

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db      37682 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 37741

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db      37742 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 37801

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db      37802 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 37861

Qy        601 ATTAAATCTTGCAACTGAA 619
              ||||||||||||||||| |
Db      37862 ATTAAATCTTGCAACTGCA 37880


RESULT 25
AAX25665
ID   AAX25665 standard; cDNA to mRNA; 7582 BP.
XX
AC   AAX25665;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Complete human endogenous retrovirus W genome.
XX
KW   Clone; human endogenous retrovirus; genome; autoimmune disease;
KW   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW   disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9902696-A1.
XX
PD   21-JAN-1999.
XX
PF   06-JUL-1998;   98WO-FR001442.
XX
PR   07-JUL-1997;   97FR-00008815.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Beseme F,  Blond J,  Bouton O,  Mandrand B,  Mallet F;
XX
DR   WPI; 1999-120897/10.
XX
PT   New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT   exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT   disease, and abnormal or failed pregnancy.
XX
PS   Claim 1; Page 71-74; 106pp; French.
```

```
XX
CC   This sequence represents the complete sequence of the human endogenous
CC   retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC   peptides encoded by them are markers of autoimmune disease (e.g. multiple
CC   sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC   insulin- dependent diabetes and related pathologies) and of abnormal or
CC   unsuccessful pregnancy and can be used as chromosomal markers for
CC   susceptibility to these conditions, or proximity markers of genes
CC   associated with this susceptibility
XX
SQ   Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 Other;

  Query Match             80.5%;  Score 511.4;  DB 2;  Length 7582;
  Best Local Similarity   89.5%;  Pred. No. 8.7e-148;
  Matches  552;  Conservative   11;  Mismatches   42;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||:||||||||||||| |||||||||||||||||| || ||||:||||||
Db       6976 CCCTGTATCTTTRACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTRAAACTA 7035

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       7036 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 7083

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||||||||:||||||| | || ||||||||||||||  ||||||||||||| |||
Db       7084 CTGGACCGGCCTGYTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 7143

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       7144 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 7203

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              : || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       7204 SGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 7263

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  |: ||||||||:|||||||||:||||||||||||
Db       7264 AGGACTAGCTGGATTTCCTAGGCTGAYTAAGAATCCYTAAGCCTAGSTGGGAAGGTGACC 7323

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| ||||||||||||||||:||||||| |||||||||||||||||
Db       7324 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGYTCACACCTGACCAATCAGAGAGCTC 7383

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||:||||||:||
Db       7384 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATYTATTGCMTG 7443

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |   ||||||||||||||||
Db       7444 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTYTTCGAGCCGGCAACGG 7503

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       7504 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 7563

Qy        601 ATTAAATCTTGCAACTG 617
              |||||||||||||:|||
Db       7564 ATTAAATCTTGCARCTG 7580


RESULT 26
AAA59215
ID   AAA59215 standard; DNA; 7582 BP.
XX
AC   AAA59215;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Human endogenous retrovirus W (HERV-W) sequence.
XX
KW   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW   gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   LTR             1. .120
FT                   /*tag=  a
FT                   /note= "R of 5' LTR"
FT   LTR             121. .575
FT                   /*tag=  b
FT                   /note= "U5 of 5' LTR"
FT   primer_bind     579. .596
FT                   /*tag=  c
FT   CDS             5581. .7194
FT                   /*tag=  d
FT                   /note= "ORF1 env538"
FT   CDS             7039. .7194
FT                   /*tag=  e
FT                   /note= "ORF2 52 AA"
FT   CDS             7112. .7255
FT                   /*tag=  f
FT                   /note= "ORF3 48 AA"
FT   misc_feature    7244. .7254
FT                   /*tag=  g
FT                   /note= "polypurine tract"
FT   LTR             7256. .7582
FT                   /*tag=  h
FT                   /note= "U3-R of 3' LTR"
FT   polyA_signal    7563. .7569
FT                   /*tag=  i
XX
PN   WO200043521-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-FR000144.
XX
PR   21-JAN-1999;   99FR-00000888.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Mallet F,  Voisset C;
XX
DR   WPI; 2000-499229/44.
XX
PT   New nucleic acid from human endogenous retrovirus, useful e.g. for
PT   diagnosis of autoimmune disease and complications of pregnancy, contains
PT   at least part of the gag gene.
XX
PS   Disclosure; Page 49-52; 53pp; French.
XX
CC   The present sequence represents an endogenetic retrovirus, which is
CC   associated with an autoimmune disease, and is integrated into the human
CC   genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
CC   HERV-W retrovirus is associated with autoimmune disease, failure of
CC   pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
CC   proteins derived from it, are useful for diagnosis of autoimmune disease
CC   (specifically multiple sclerosis) and for monitoring pregnancy. The
CC   nucleic acid fragments may also be used for in situ labelling of isolated
CC   chromosomes, while the transcription product can be used to study or
CC   monitor T cell proliferation in vitro
XX
SQ   Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 0 U; 216 Other;

  Query Match             80.5%;  Score 511.4;  DB 3;  Length 7582;
  Best Local Similarity   89.5%;  Pred. No. 8.7e-148;
  Matches  552;  Conservative   11;  Mismatches   42;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||:||||||||||||| |||||||||||||||||| || ||||:||||||
Db       6976 CCCTGTATCTTTRACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTRAAACTA 7035

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       7036 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 7083

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||||||||:||||||| | || ||||||||||||||  ||||||||||||| |||
Db       7084 CTGGACCGGCCTGYTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 7143

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       7144 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 7203

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              : || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       7204 SGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 7263

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  |: ||||||||:|||||||||:||||||||||||
Db       7264 AGGACTAGCTGGATTTCCTAGGCTGAYTAAGAATCCYTAAGCCTAGSTGGGAAGGTGACC 7323

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| ||||||||||||||||:||||||| |||||||||||||||||
Db       7324 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGYTCACACCTGACCAATCAGAGAGCTC 7383

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||:||||||:||
Db       7384 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATYTATTGCMTG 7443

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       7444 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTYTTCGAGCCGGCAACGG 7503

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       7504 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 7563

Qy        601 ATTAAATCTTGCAACTG 617
              |||||||||||||:|||
Db       7564 ATTAAATCTTGCARCTG 7580


RESULT 27
ABN97978
ID   ABN97978 standard; DNA; 46340 BP.
XX
AC   ABN97978;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral sequence HI3.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 15; Page 186-199; 225pp; French.
```

```
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 46340 BP; 16104 A; 8738 C; 8434 G; 13064 T; 0 U; 0 Other;

  Query Match             80.4%;  Score 510.6;  DB 3;  Length 46340;
  Best Local Similarity   90.0%;  Pred. No. 3.7e-147;
  Matches  575;  Conservative    0;  Mismatches   49;  Indels   15;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||  |||||| |||
Db      33280 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGTTGTAAAGCTA 33339

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||| | ||||||| ||||||||||||||||| |||||||||||
Db      33340 CAAATCGTTCTTCAAATGGAACCCCAGATGAAGTCCATGACTAAGATCTACCGTGGACCC 33399

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||||||| |||||||||||||| ||  |||||| || ||| ||||||||||||||
Db      33400 CTGGACCGGCCTACTAGCCCATGCTCCAATTGTAATGATATCGAACGCACCCCTCCCGAG 33459

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||||||||||||| || |||||||||||||  |||
Db      33460 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCCGCAGGAAGCAGTTAGACTGGT 33519

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      33520 CGTCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGTGGGGGGACTGAGAGAC 33579

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||| ||||||||||||||||||| || ||||||||| |||||||||||||||||||||
Db      33580 AGGATTAGCTGGATTTCCTAGGCCGACTAAGAATCCCAAAGCCTAGCTGGGAAGGTGACC 33639

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
               ||||||||| ||||||  ||||||||||||||||||||||||||||||||
Db      33640 ACATCCACCTTTAAACACTGGGCTTGCAACTTAGCTCACACCCGACCAATCAGGTAGTAA 33699

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              |||||||||||||||||||||||||| |||||| |||||||||| |||||||||||||||
Db      33700 AGAGAGCTCACTAAAATGCTAATTAGACAAAAACAGGAGGTAAAAAAATAGCCAATCATC 33759

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              ||| ||||||||||||||||||| |||||| ||||||||||||||||||||||||| |||
Db      33760 TATCGCCTGAGAGCACAGCGGGAAGGACAATGATCGGGATATAAACCCAGGCATTCAAGC 33819

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              |||||||||| |||  |||||||||||||||||||||||||| ||||| | |
Db      33820 CGGCAACGGCTACCTTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTCTCT------GT 33873

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAA 630
               ||||||||||||||| ||||||||| |||||||||| |
Db      33874 CTTCACTCTATTAAATATTGCAACTGCAAAAAAAAAATA 33912


RESULT 28
ABN97929
ID   ABN97929 standard; DNA; 10499 BP.
XX
AC   ABN97929;
XX
DT   01-AUG-2002  (first entry)
```

```
XX
DE   Human retroviral sequence HERV-7q.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis; treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 3; Fig 1; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 U; 0 Other;

  Query Match             79.7%;  Score 505.8;  DB 3;  Length 10499;
  Best Local Similarity   91.1%;  Pred. No. 5.6e-146;
  Matches  564;  Conservative    0;  Mismatches   42;  Indels   13;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       9274 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 9333

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       9334 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 9381

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       9382 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 9441

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       9442 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 9501

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       9502 C-TCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 9560

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       9561 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 9620

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       9621 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 9680

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db       9681 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 9740

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       9741 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 9800

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       9801 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 9860

Qy        601 ATTAAATCTTGCAACTGAA 619
              ||||||||||||||||| |
Db       9861 ATTAAATCTTGCAACTGCA 9879


RESULT 29
AAI14608
ID   AAI14608 standard; DNA; 1894 BP.
XX
AC   AAI14608;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #4541 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 4541; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;
```

```
  Query Match             78.7%;  Score 499.6;  DB 4;  Length 1894;
  Best Local Similarity   88.7%;  Pred. No. 2.1e-144;
  Matches  571;  Conservative    0;  Mismatches   54;  Indels   19;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||| |||||||||| || | |||||||||
Db       1217 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAATCGAAGCAGTAAAACTA 1276

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||| ||| ||| |||  ||||||
Db       1277 CAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAGTAAAATCTACCACGGACCC 1336

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||||||||||||||||||||| ||||||||||||||  |||||||||||||||||
Db       1337 CTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAAAGGCACCCCTCCCGAG 1396

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||| |||||| ||||||||||||| |||||||||||||| |||
Db       1397 GAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGT 1456

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
                |  ||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       1457 TGTTGGCCAACCTCCCCAACAGCAGTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 1516

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||| || || |||||||||| ||||| ||||||||||||| |||||||||||||||||
Db       1517 AGGAATAACTAGATTTCCTAGACCAACTAAGAATCCCTAAGACTAGCTGGGAAGGTGACC 1576

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
              || ||||||| ||||||  |||||||||||||||||||| ||| |||||||
Db       1577 GCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCAACCAATCAGATACTAA 1636

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              ||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||||||
Db       1637 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGATAAAGAAATAGCCAATCATC 1696

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              | |||||||| |||||||| |||||||||| |||||||||||||||||||||||||||||
Db       1697 TGTTGCCTGACAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCATTCGAGC 1756

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              | || || || |||| |||||||||||||||||||||||||| |||||||
Db       1757 CAGCTACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT---------- 1806

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
               ||||||||||||||||||||||||| ||||| ||||| | |||
Db       1807 CTTCACTCTATTAAATCTTGCAACTGCAAAAATAAAAATAGAAA 1850

RESULT 30
ABA56337
ID   ABA56337 standard; DNA; 1894 BP.
XX
AC   ABA56337;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #4642.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
```

```
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 1; SEQ ID NO 4642; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

  Query Match             78.7%;  Score 499.6;  DB 4;  Length 1894;
  Best Local Similarity   88.7%;  Pred. No. 2.1e-144;
  Matches  571;  Conservative    0;  Mismatches   54;  Indels   19;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||| |||||||||| || | |||||||||
Db       1217 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAATCGAAGCAGTAAAACTA 1276

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||| ||| ||| |||  ||||||
Db       1277 CAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAGTAAAATCTACCACGGACCC 1336

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||||||||||||||||||||| ||||||||||||||  |||||||||||||||||
Db       1337 CTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAAAGGCACCCCTCCCGAG 1396

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||| |||||| ||||||||||||| |||||||||||||| |||
Db       1397 GAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGT 1456

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
                |  ||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       1457 TGTTGGCCAACCTCCCCAACAGCAGTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 1516

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||| || || |||||||||| ||||| ||||||||||||| |||||||||||||||||
Db       1517 AGGAATAACTAGATTTCCTAGACCAACTAAGAATCCCTAAGACTAGCTGGGAAGGTGACC 1576

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
              || ||||||| ||||||  |||||||||||||||||||| ||| |||||||
Db       1577 GCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCAACCAATCAGATACTAA 1636

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              ||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||||||
Db       1637 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGATAAAGAAATAGCCAATCATC 1696

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              | |||||||| |||||||| |||||||||| |||||||||||||||||||||||||||||
Db       1697 TGTTGCCTGACAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCATTCGAGC 1756

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              | || || || |||| |||||||||||||||||||||||||| |||||||
Db       1757 CAGCTACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT---------- 1806

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
               ||||||||||||||||||||||||| ||||| ||||| | |||
Db       1807 CTTCACTCTATTAAATCTTGCAACTGCAAAAATAAAAATAGAAA 1850
```

```
Search completed: January 22, 2005, 16:40:17
Job time : 302.418 secs
```

```
                      GenCore version 5.1.6
             Copyright (c) 1993 - 2005  Compugen Ltd.


OM nucleic - nucleic search, using sw model
                -
Run on:          January 22, 2005, 12:03:14 ; Search time 58.7997 Seconds
                                           (without alignments)
                                           7676.074 Million cell updates/sec

Title:           US-09-319-156B-6
Perfect score:   635
Sequence:        1 ccctgtatctttaacctcct..........tgaaaaaaaaaaaaaaaaa 635

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters:      1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                 2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                 3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                 4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                 5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                 6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                                  SUMMARIES
                   %
Result           Query
  No.    Score   Match Length DB  ID                         Description
 -----------------------------------------------------------------------------
    1      635   100.0    635  4  US-08-979-847B-102          Sequence 102, App
    2      545    85.8   1329  4  US-08-979-847B-108          Sequence 108, App
    3    533.8    84.1   2946  3  US-09-175-928-3             Sequence 3, Appli
    4    233.6    36.8    279  1  US-08-686-878A-50           Sequence 50, Appl
    5    233.6    36.8    279  1  US-08-721-489-4             Sequence 4, Appli
c   6    196.6    31.0    456  4  US-09-621-976-9366          Sequence 9366, Ap
    7      128    20.2   1859  3  US-08-691-563C-46           Sequence 46, Appl
    8      128    20.2   1859  4  US-09-374-766-46            Sequence 46, Appl
    9      128    20.2   1859  4  US-08-979-847B-42           Sequence 42, Appl
   10    125.8    19.8  80246  3  US-09-078-294-4             Sequence 4, Appli
   11    125.8    19.8  80595  3  US-09-078-294-3             Sequence 3, Appli
   12      125    19.7   1722  3  US-08-691-563C-58           Sequence 58, Appl
   13      125    19.7   1722  4  US-09-374-766-58            Sequence 58, Appl
   14      125    19.7   1722  4  US-08-979-847B-54           Sequence 54, Appl
   15    110.6    17.4    149  4  US-09-513-999C-24228        Sequence 24228, A
   16     76.4    12.0   3910  3  US-09-120-653D-1            Sequence 1, Appli
   17     73.4    11.6    492  4  US-09-621-976-11057         Sequence 11057, A
   18     71.8    11.3    443  4  US-09-513-999C-21394        Sequence 21394, A
   19     68.4    10.8   1481  4  US-08-979-847B-105          Sequence 105, App
   20     41.4     6.5    342  4  US-09-513-999C-13878        Sequence 13878, A
   21     41.2     6.5   3824  4  US-10-140-002-541           Sequence 541, App
   22     40.4     6.4    464  3  US-09-120-653D-6            Sequence 6, Appli
c  23     38.6     6.1     47  4  US-09-422-978-3579          Sequence 3579, Ap
c  24     37.6     5.9   7218  1  US-08-232-463-14            Sequence 14, Appl
   25       36     5.7    396  4  US-09-854-133-539           Sequence 539, App
   26     35.2     5.5   1497  4  US-09-220-132-94            Sequence 94, Appl
c  27       34     5.4   3568  3  US-09-218-363-3             Sequence 3, Appli
```

```
   28     33.8     5.3    117  4  US-09-621-976-9850          Sequence 9850, Ap
   29     33.8     5.3    177  4  US-09-621-976-1047          Sequence 1047, Ap
c  30       33     5.2   2417  1  US-08-011-398B-1            Sequence 1, Appli
c  31       33     5.2   2417  1  US-08-464-051-1             Sequence 1, Appli
c  32       33     5.2   2417  2  US-08-462-498-1             Sequence 1, Appli
c  33       33     5.2   2417  3  US-08-554-385-2             Sequence 2, Appli
   34     32.8     5.2    296  4  US-09-513-999C-24392        Sequence 24392, A
c  35     32.8     5.2   1446  4  US-09-149-476-139           Sequence 139, App
   36     32.6     5.1    598  3  US-09-247-155-42            Sequence 42, Appl
   37     32.6     5.1   2006  4  US-09-354-123-5             Sequence 5, Appli
c  38     32.6     5.1  90541  4  US-09-759-359A-3            Sequence 3, Appli
c  39     32.6     5.1  90541  4  US-10-207-973-3             Sequence 3, Appli
   40     32.4     5.1    427  4  US-09-702-705-854           Sequence 854, App
   41     32.4     5.1    427  4  US-09-736-457-854           Sequence 854, App
   42     32.4     5.1    427  4  US-09-614-124B-854          Sequence 854, App
   43     32.4     5.1    427  4  US-09-671-325-854           Sequence 854, App
   44     32.4     5.1    427  4  US-09-658-824-854           Sequence 854, App
   45       32     5.0   2494  3  US-09-724-864-6             Sequence 6, Appli

                                ALIGNMENTS


RESULT 1
US-08-979-847B-102
; Sequence 102, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLA
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,847B
;            FILING DATE: 26-No. 6582703-1997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;  INFORMATION FOR SEQ ID NO: 102:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 635 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-08-979-847B-102
```

```
 Query Match             100.0%;  Score 635;  DB 4;  Length 635;
 Best Local Similarity   100.0%;  Pred. No. 3.5e-202;
 Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy       61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy      121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy      181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy      241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy      301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy      361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy      421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy      481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy      541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy      601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
            |||||||||||||||||||||||||||||||||||
Db      601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635


RESULT 2
US-08-979-847B-108
; Sequence 108, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
;         APPLICANT: PERRON, HERVE
;                    BESEME, FREDERIC
;                    BEDIN, FREDERIC
;                    PARANHOS-BACCALA, GLAUCIA
;                    KOMURIAN-PRADEL, FLORENCE
;                    JOLIVET-REYNAUD, COLETTE
;                    MANDRAND, BERNARD
;                    GARSON, JEREMY
;                    TUKE, PHILIP
;         TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                             ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                             THERAPEUTIC PURPOSES
;         NUMBER OF SEQUENCES: 210
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: OLIFF & BERRIDGE, PLC
;              STREET: P.O. BOX 19928
;              CITY: ALEXANDRIA
;              STATE: VA
;              COUNTRY: USA
;              ZIP: 22320
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/08/979,847B
;              FILING DATE: 26-No. 6582703-1997
;              CLASSIFICATION: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: BERRIDGE, WILLIAM P.
;              REGISTRATION NUMBER: 30,024
;              REFERENCE/DOCKET NUMBER: WPB 39046A
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 703-836-6400
;              TELEFAX: 703-836-2787
;    INFORMATION FOR SEQ ID NO: 108:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1329 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 108:
US-08-979-847B-108

 Query Match             85.8%;  Score 545;  DB 4;  Length 1329;
 Best Local Similarity   92.4%;  Pred. No. 7.3e-172;
 Matches  572;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
            ||||||||||| ||| ||||||||||||||||||||||||||||  || ||||||| |||
Db      257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316

Qy       61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
            ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db      317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376

Qy      121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
            ||||||||||||||||| | |||||| |||||||||||||||||||| ||||||||||||
Db      377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436

Qy      181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
            |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db      437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496

Qy      241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db      497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556

Qy      301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
            |||||||||||||||||||||||  || ||||||||| |||||||| |||||||||||||
Db      557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616

Qy      361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
            ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db      617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676

Qy      421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
            |||||||||||||| ||||||||| |||||||||||| ||||||||||||||||||||||
Db      677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736

Qy      481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
            |||||||||||||| |||||||||| |||||||||||| ||||||||| |||| ||||| |
Db      737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796

Qy      541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
            ||||||||||||||||||||||| |||||||||| |||||||||||||||||||||||||
Db      797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856

Qy      601 ATTAAATCTTGCAACTGAA 619
```

```
              |||||||| |||||||| |
Db        857 ATTAAATCATGCAACTGCA 875


RESULT 3
US-09-175-928-3
; Sequence 3, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT: McCoy, John M.
;  APPLICANT: LaVallie, Edward R.
;  APPLICANT: Collins-Racie, Lisa A.
;  APPLICANT: Evans, Cheryl
;  APPLICANT: Merberg, David
;  APPLICANT: Treacy, Maurice
;  APPLICANT: Mi, Sha
;  APPLICANT: Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6006B.AJ172A
;  CURRENT APPLICATION NUMBER: US/09/175,928A
;  CURRENT FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-175-928-3

  Query Match             84.1%;  Score 533.8;  DB 3;  Length 2946;
  Best Local Similarity   91.5%;  Pred. No. 6.8e-168;
  Matches  581;  Conservative    0;  Mismatches   42;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2323 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2382

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2383 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2430

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2431 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2490

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2491 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2550

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2551 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2610

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||| 
Db       2611 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2670

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2671 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2730

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       2731 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2790

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2791 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2850

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2851 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2910

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       2911 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2945


RESULT 4
US-08-686-878A-50
; Sequence 50, Application US/08686878A
; Patent No. 5708157
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Evans, Cheryl
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/686,878A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  50:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  279 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-686-878A-50

  Query Match             36.8%;  Score 233.6;  DB 1;  Length 279;
  Best Local Similarity   88.1%;  Pred. No. 4.1e-68;
  Matches  245;  Conservative    1;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        358 ACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAG 417
              :|  ||||||||| |||||| |||| ||||||   || | ||||  ||||||||||||||
Db          1 RCCACATCCACCTTTAAACACGGGGNTTGCAAANAAGATNACACTTGACCAATCAGAGAG 60

Qy        418 CTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGC 477
               ||| ||||||| | ||| ||||||| |||||||||||||||||||||||||||||||||
Db         61 NTCANTAAAATGATNATTNGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGC 120

Qy        478 CTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAA 537
              ||||||||||||| |||||||||| |||||||||||||||||| |  || ||||||||||
Db        121 CTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTNGAGCCGGCAA 180

Qy        538 CGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCAC 597
              |||||||||||||||||||||||||||||||||||| ||| ||||||||  |||||||| 
Db        181 CGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTNTGTTTTCATGCTATTTCAN 240
```

```
Qy        598 TCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              | |||||||| ||||||||| |||||||||||||||||
Db        241 TNTATTAAATNTTGCAACTGCAAAAAAAAAAAAAAAAA 278

RESULT 5
US-08-721-489-4
; Sequence 4, Application US/08721489
; Patent No. 5786465
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/721,489
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  279 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-721-489-4

  Query Match             36.8%;  Score 233.6;  DB 1;  Length 279;
  Best Local Similarity   88.1%;  Pred. No. 4.1e-68;
  Matches  245;  Conservative    1;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        358 ACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAG 417
              :|  ||||||||| |||||| |||| ||||||   || | ||||  ||||||||||||||
Db          1 RCCACATCCACCTTTAAACACGGGGNTTGCAAANAAGATNACACTTGACCAATCAGAGAG 60

Qy        418 CTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGC 477
               ||| ||||||| | ||| |||||||| ||||||||||||||||||||||||||||||||
Db         61 NTCANTAAAATGATNATTNGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGC 120

Qy        478 CTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAA 537
              ||||||||||||| |||||||||| |||||||||||||||||| |  || ||||||||||
Db        121 CTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTNGAGCCGGCAA 180

Qy        538 CGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCAC 597
              |||||||||||||||||||||||||||||||||||| ||| ||||||||  ||||||||
Db        181 CGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTNTGTTTTCATGCTATTTCAN 240

Qy        598 TCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              | |||||||| ||||||||| |||||||||||||||||
```

```
Db        241 TNTATTAAATNTTGCAACTGCAAAAAAAAAAAAAAAAA 278

RESULT 6
US-09-621-976-9366/c
; Sequence 9366, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 9366
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-9366

  Query Match             31.0%;  Score 196.6;  DB 4;  Length 456;
  Best Local Similarity   84.9%;  Pred. No. 1.3e-55;
  Matches  264;  Conservative    0;  Mismatches   24;  Indels   23;  Gaps    3;

Qy        292 CTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGG 351
              ||||||||||||||||||||||||||||||||| || |||||||||||||||||||||||
Db        307 CTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGG 248

Qy        352 AAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC 411
              ||||||||  ||||||||| ||| || |||||||||||||||||||||||||||||| ||
Db        247 AAGGTGACCACATCCACCTTTAAGCAGGGGGCTTGCAACTTAGCTCACACCCGACCAGTC 188

Qy        412 ---------AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAG 462
                       || ||||||||||||||||||||||||| |||| ||||||||||  |||||
Db        187 AGGTAGTAAAGCGAGCTCACTAAAATGCTAATTAGGCTAAAACAGGAGGTAAA-CAATAG 129

Qy        463 CCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGG 522
              ||||||||||||  ||| |||||||||| |||||||||| ||| |||||||||||
Db        128 CCAATCATCTATCTCCTCAGAGCACAGCAGGAGGGACAATGATAGGGATATAAAC----- 74

Qy        523 CATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTT 582
                      | |||||||||  ||| || ||||||||||||||||||||||| ||||||||
Db         73 --------CTGGCAACGGCTGCCCTCTCTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTT 22

Qy        583 TTCACTCTATT 593
              |||||||||||
Db         21 TTCACTCTATT 11


RESULT 7
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563C
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-691-563C-46

  Query Match             20.2%;  Score 128;  DB 3;  Length 1859;
  Best Local Similarity   62.7%;  Pred. No. 3e-32;
  Matches  237;  Conservative    0;  Mismatches  130;  Indels   11;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||| |||||||||||||| || ||||| ||||| || ||| |  |  | || |||
Db       1446 CCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCTA 1505

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || ||  | |||||||| | ||| |||   || |  |  ||  || || | ||||||
Db       1506 CAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACCC 1565

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              || ||||  ||  || ||||         | | | || | |  |  |  ||||||  |||
Db       1566 CTAGACCAACCCCCTGGCCC---------TTTCACTGGCCTAAAGAGTTCCCCTCTGGAG 1616

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              || |   | ||||||   ||||  || |||||| || |||  ||||| || |||||| ||
Db       1617 GACACTACCACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGT 1676

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||   ||||  | ||||| ||||  |||| | ||| ||| |||| ||||| ||||||
Db       1677 CATTGCCCAA--TTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGT 1734

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
                  | ||||||| |||   | |    |  || |     | | |   ||   || | |
Db       1735 GAAGCCAGCTGGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAG 1794

Qy        361 GCATCCACCTCTAAACAT 378
              |  |  |  |  ||  ||
Db       1795 GATTGTAAATGCAACAAT 1812

RESULT 8
US-09-374-766-46
; Sequence 46, Application US/09374766
; Patent No. 6579526
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
```

```
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/374,766
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-374-766-46

  Query Match             20.2%;  Score 128;  DB 4;  Length 1859;
  Best Local Similarity   62.7%;  Pred. No. 3e-32;
  Matches  237;  Conservative    0;  Mismatches  130;  Indels   11;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||| |||||||||||||| || ||||| ||||| || ||| |  |  | || |||
Db       1446 CCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCTA 1505

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || ||  | |||||||| | ||| |||   || |  |  ||  || || | ||||||
Db       1506 CAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACCC 1565

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              || ||||  ||  || ||||         | | | || | |  |  |  ||||||  |||
Db       1566 CTAGACCAACCCCCTGGCCC---------TTTCACTGGCCTAAAGAGTTCCCCTCTGGAG 1616

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              || |   | ||||||   ||||  || |||||| || |||  ||||| || |||||| ||
Db       1617 GACACTACCACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGT 1676

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||   ||||  | ||||| ||||  |||| | ||| ||| |||| ||||| ||||||
Db       1677 CATTGCCCAA--TTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGT 1734

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
                  | ||||||| |||   | |    |  || |     | | |   ||   || | |
Db       1735 GAAGCCAGCTGGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAG 1794

Qy        361 GCATCCACCTCTAAACAT 378
              |  |  |  |  ||  ||
Db       1795 GATTGTAAATGCAACAAT 1812

RESULT 9
US-08-979-847B-42
; Sequence 42, Application US/08979847B
; Patent No. 6582703
;    GENERAL INFORMATION:
;         APPLICANT: PERRON, HERVE
;                    BESEME, FREDERIC
;                    BEDIN, FREDERIC
;                    PARANHOS-BACCALA, GLAUCIA
;                    KOMURIAN-PRADEL, FLORENCE
;                    JOLIVET-REYNAUD, COLETTE
;                    MANDRAND, BERNARD
;                    GARSON, JEREMY
;                    TUKE, PHILIP
;         TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                             ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                             THERAPEUTIC PURPOSES
;         NUMBER OF SEQUENCES: 210
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: OLIFF & BERRIDGE, PLC
;              STREET: P.O. BOX 19928
;              CITY: ALEXANDRIA
;              STATE: VA
;              COUNTRY: USA
;              ZIP: 22320
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/08/979,847B
;              FILING DATE: 26-No. 6582703-1997
;              CLASSIFICATION: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: BERRIDGE, WILLIAM P.
;              REGISTRATION NUMBER: 30,024
;              REFERENCE/DOCKET NUMBER: WPB 39046A
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 703-836-6400
;              TELEFAX: 703-836-2787
;    INFORMATION FOR SEQ ID NO: 42:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1859 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-08-979-847B-42

  Query Match             20.2%;  Score 128;  DB 4;  Length 1859;
  Best Local Similarity   62.7%;  Pred. No. 3e-32;
  Matches  237;  Conservative    0;  Mismatches  130;  Indels   11;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||| |||||||||||||| || ||||| ||||| || |||  |  |  | || |||
Db       1446 CCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCTA 1505

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || ||  | |||||||| | ||| |||   || |  |  ||  || || | ||||||
Db       1506 CAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAACTTCTACTGAGGACCC 1565

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              || ||||  ||  || ||||         | | | || | |  |  |  ||||||  |||
Db       1566 CTAGACCAACCCCCTGGCCC---------TTTCACTGGCCTAAAGAGTTCCCCTCTGGAG 1616

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              || |   | ||||||   ||||  || |||||| || |||  ||||| || |||||| ||
Db       1617 GACACTACCACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGAAGTAGCTAGAGCAGT 1676

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||   ||||  | ||||| ||||  |||| | ||| ||| |||| ||||| ||||||
Db       1677 CATTGCCCAA--TTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGGATTGAGAGGT 1734

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
                  | ||||||| |||   | |    |  || |     | | |   ||   || | |
```

```
Db       1735 GAAGCCAGCTGGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTTTGTGTCTAGCTAAAG 1794

Qy        361 GCATCCACCTCTAAACAT 378
              |  |  |  |  ||  ||
Db       1795 GATTGTAAATGCAACAAT 1812


RESULT 10
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

  Query Match             19.8%;  Score 125.8;  DB 3;  Length 80246;
  Best Local Similarity   67.2%;  Pred. No. 1.8e-30;
  Matches  213;  Conservative    0;  Mismatches   92;  Indels   12;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||| |||||||||| || || ||||| || || || || ||| |  |  | || |||
Db      58888 CCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCTA 58947

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || ||  |  | |||||| | | ||| |||   || |  |  ||  ||  | | ||||||
Db      58948 CAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACCC 59007

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||| || ||| ||||         | | |||| | |  |  || ||||||  |||
Db      59008 CTGGACCGACCCGCTGGCCC---------TTTCAATGGCCTAAAGAGCTCCCCTCTGGAG 59058

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              || |   | ||||||   ||||| ||   |||| || ||||| ||||| || || ||||||
Db      59059 GACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGT 59118

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||  | ||    ||||||||||  |||| | ||||||| ||||||||| ||||||
Db      59119 CATCGCCAAA---TCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGT 59175

Qy        301 AGGACTAGCTGGATTTC 317
                  | ||||||  |||
Db      59176 GAAGCCAGCTGGGCTTC 59192


RESULT 11
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 80595
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

  Query Match             19.8%;  Score 125.8;  DB 3;  Length 80595;
  Best Local Similarity   67.2%;  Pred. No. 1.8e-30;
  Matches  213;  Conservative    0;  Mismatches   92;  Indels   12;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||| ||||||||| | || || ||||| || || || ||| |  |  | || |||
Db      59121 CCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATCGAGGCCATCAAGCTA 59180

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || ||  |  | |||||| | | ||| |||   || |  |  ||  ||  | | ||||||
Db      59181 CAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAACTTCTGCTGAGGACCC 59240

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||| || ||| ||||         | | |||| | |  |  || ||||||  |||
Db      59241 CTGGACCGACCCGCTGGCCC---------TTTCAATGGCCTAAAGAGCTCCCCTCTGGAG 59291

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              || |   | ||||||   ||||| ||   |||| || |||| ||||| || || ||||||
Db      59292 GACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGAAGTAGCTACAGCGGT 59351

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||  | ||     ||||||||||  |||| | |||||||  ||||||||| ||||||
Db      59352 CATCGCCAAA---TCCCAACAGCAGCTGGGGTGTCCTGTTTGGAGGGGGGATTGAGAGGT 59408

Qy        301 AGGACTAGCTGGATTTC 317
                  | ||||||  |||
Db      59409 GAAGCCAGCTGGGCTTC 59425


RESULT 12
US-08-691-563C-58
; Sequence 58, Application US/08691563C
; Patent No. 6001987
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563C
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 base pairs
```

```
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-691-563C-58

  Query Match             19.7%;  Score 125;  DB 3;  Length 1722;
  Best Local Similarity   69.2%;  Pred. No. 2.8e-31;
  Matches  204;  Conservative    0;  Mismatches   80;  Indels   11;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||| ||||| || ||||| || ||||| || || || ||| || |  | || |||
Db        796 CCCTGTATTTTTAAGCTTCTTGTCAAATTTGTTTCCTCTAGGATCGAAGCCATCAAGCTA 855

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || ||  | |||||||| | ||| |||   || |  |  ||  || |||  ||||||
Db        856 CAGATGGTCTTACAAATGGAACCCCAAATGAGTTCAACTAACAACTTCTACCAAGGACCC 915

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||| ||  |  || || | | |          |  |   |  ||   ||||||  ||
Db        916 CTGGAACGATCCACTGGCACTTCC----------ACTAGCCTAGAGATTCCCCTCTGGAA 965

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              || |   ||||||||   ||||| || |||||| || |||| ||||| || |||||||||
Db        966 GACACTACAACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGCGGT 1025

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGA 295
              |||| |||||  | |||||||||| ||||| | ||||||| |||||||||| |||
Db       1026 CATCGGCCAA-ATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGA 1079


RESULT 13
US-09-374-766-58
; Sequence 58, Application US/09374766
; Patent No. 6579526
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/374,766
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  58:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 base pairs
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-374-766-58

  Query Match             19.7%;  Score 125;  DB 4;  Length 1722;
  Best Local Similarity   69.2%;  Pred. No. 2.8e-31;
  Matches  204;  Conservative    0;  Mismatches   80;  Indels   11;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||| ||||| || ||||| || ||||| || || || ||| || |  | || |||
Db        796 CCCTGTATTTTTAAGCTTCTTGTCAAATTTGTTTCCTCTAGGATCGAAGCCATCAAGCTA 855

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || ||  | |||||||| | ||| |||   || |  |  ||  || |||  ||||||
Db        856 CAGATGGTCTTACAAATGGAACCCCAAATGAGTTCAACTAACAACTTCTACCAAGGACCC 915

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||| ||  |  || || | | |          |  |   |  ||   ||||||  ||
Db        916 CTGGAACGATCCACTGGCACTTCC----------ACTAGCCTAGAGATTCCCCTCTGGAA 965

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              || |   ||||||||   ||||| || |||||| || |||| ||||| || |||||||||
Db        966 GACACTACAACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGCGGT 1025

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGA 295
              |||| |||||  | |||||||||| ||||| | ||||||| |||||||||| |||
Db       1026 CATCGGCCAA-ATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGA 1079


RESULT 14
US-08-979-847B-54
; Sequence 54, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
```

```
;             TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 54:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1722 base pairs
;             TYPE: nucleotide
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-08-979-847B-54

  Query Match             19.7%;  Score 125;  DB 4;  Length 1722;
  Best Local Similarity   69.2%;  Pred. No. 2.8e-31;
  Matches  204;  Conservative    0;  Mismatches   80;  Indels   11;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||| ||||| || ||||| || ||||| || || || ||| || |  | || |||
Db        796 CCCTGTATTTTTAAGCTTCTTGTCAAATTTGTTTCCTCTAGGATCGAAGCCATCAAGCTA 855

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || ||  | |||||||| | ||| |||   || |  |  ||  || |||  ||||||
Db        856 CAGATGGTCTTACAAATGGAACCCCAAATGAGTTCAACTAACAACTTCTACCAAGGACCC 915

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||| ||  |  || || | | |          |  |   |  ||   ||||||  ||
Db        916 CTGGAACGATCCACTGGCACTTCC----------ACTAGCCTAGAGATTCCCCTCTGGAA 965

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              || |   ||||||||   ||||| || |||||| || |||| ||||| || |||||||||
Db        966 GACACTACAACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGAAGTAGCTAGAGCGGT 1025

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGA 295
              |||| |||||  | |||||||||| ||||| | ||||||| |||||||||| |||
Db       1026 CATCGGCCAA-ATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAGGGGGGATTGA 1079


RESULT 15
US-09-513-999C-24228
; Sequence 24228, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 24228
;   LENGTH: 149
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 127
;   OTHER INFORMATION: r=a or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 129
;   OTHER INFORMATION: n=a, g, c or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 135
;   OTHER INFORMATION: m=a or c
US-09-513-999C-24228

  Query Match             17.4%;  Score 110.6;  DB 4;  Length 149;
  Best Local Similarity   90.1%;  Pred. No. 3.9e-27;
  Matches  127;  Conservative    2;  Mismatches   11;  Indels    1;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              | |||||||||||||||||||||||||||||||||||||||||||||| || ||||||||
Db          5 CTCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTATAAAACTA 64

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| ||||| |||||||||| ||||||| ||||||||||||||||| || | ||||| 
Db         65 CAAATGGTTCTCCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACTGCGGACCT 124

Qy        121 CTGG-ACCGGCCTGCTAGCCC 140
              ||:| |||||:||||||||||
Db        125 CTRGNACCGGMCTGCTAGCCC 145


RESULT 16
US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
;  APPLICANT: YOON, JI-WON
;  APPLICANT:  JUN, HEE-SOOK
;  APPLICANT:  PARK, HAE-JOON
;  APPLICANT:  AHN, JONG-SEONG
;  APPLICANT:  HA, YOUNG-JU
;  APPLICANT:  CHUNG, SOO-IL
;  TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
;  FILE REFERENCE: 98-338
;  CURRENT APPLICATION NUMBER: US/09/120,653D
;  CURRENT FILING DATE:  1998-07-22
;  PRIOR APPLICATION NUMBER: KR 98-10108
;  PRIOR FILING DATE: 1998-03-24
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 1
;   LENGTH: 3910
;   TYPE: DNA
;   ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

  Query Match             12.0%;  Score 76.4;  DB 3;  Length 3910;
  Best Local Similarity   61.4%;  Pred. No. 8.8e-15;
  Matches  178;  Conservative    0;  Mismatches  101;  Indels   11;  Gaps    3;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||| | || |||||| | ||||| || ||||| |  || ||| || |  |  | || |||
Db       3194 CCCAGCATTTTTAACGTTCTTGTCAAATTTGT-TTGTCTAGATTCGAGGCCTTCAAGCTA 3252

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || ||  | || ||  | | ||  ||    || |  |   |  || || | |||| |
Db       3253 CAGATGGTCTTACATATCAAACCCCGAATAATTTCAACTAACTACTTCTACGGAGGACAC 3312

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||   || ||| ||   | | |         || | | ||  |  ||||||   ||
Db       3313 CTGGACTAACCAGCTGGCAGTTACCC---------TGGCCTAGAGAGTTCCCCTCTGAAG 3363

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |  |   |||||||| | |||||  |   | || || || | ||| | |  |||| | ||
Db       3364 GTCACTACAACTGCAAAGCCCCTTTTTCGCGCCTATCCATCAGGACGTACCTAGAACAGT 3423

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGG 290
              | || ||| |  |  ||||||||| ||||  | |||||||||  | ||||
Db       3424 CCTCGGCC-ATATTGCCAACAGCAGTTGGAGTGTCCTGTTGATTGAGGGG 3472


RESULT 17
US-09-621-976-11057
; Sequence 11057, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
```

```
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 11057
;   LENGTH: 492
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-11057

  Query Match             11.6%;  Score 73.4;  DB 4;  Length 492;
  Best Local Similarity   60.1%;  Pred. No. 2.3e-14;
  Matches  164;  Conservative    0;  Mismatches   96;  Indels   13;  Gaps    2;

Qy         24 TAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATGGAGCA 83
              |||    || |  || || |||||  |  | || || ||||| |   | |||||||| | 
Db         95 TAATCAAGTGTAATCAAGGATCAAGGCCATCAAGCTGCAAATGGACTTACAAATGGAACC 154

Qy         84 CCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAGCCCATG 143
               || |||   || |         ||  | | | ||||||||| |  |||||| | || | 
Db        155 TCAAATGAGCTCAACTCACGGCTTCTACTGAGAACCCCTGGATCCACCTGCTGGTCCCTC 214

Qy        144 CTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCT 203
                        | | | |  || |  ||||||  ||||| | | ||||||||   |||||
Db        215 ----------ACTAGCCTAAAAAGTTCCCCTCTGGAGGACACCACAACTGCAGGGCCCCT 264

Qy        204 ACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGC 263
               ||   |||| |  || | ||||| ||  ||| ||   |          | |||||||||
Db        265 TCTTCACCCCTAACCATCAGGAAGTAGCCAGAACG---ACTGCTGCCAGTTCCCAACAGC 321

Qy        264 ACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAG 296
              | ||||  | ||||||  |||||||||||||||
Db        322 AATTGGTGTGTCCTGTCTAGAGGGGGGACTGAG 354


RESULT 18
US-09-513-999C-21394
; Sequence 21394, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 21394
;   LENGTH: 443
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 102
;   OTHER INFORMATION: k=g or t
US-09-513-999C-21394

  Query Match             11.3%;  Score 71.8;  DB 4;  Length 443;
  Best Local Similarity   89.2%;  Pred. No. 7.5e-14;
  Matches   99;  Conservative    1;  Mismatches    8;  Indels    3;  Gaps    2;

Qy        510 ATATAAACCCAGGCATTCGAGCCGGCAACGGCA-ACCCCCTTTGGGTCCCCTCCCTTTGT 568
              |||||||||||||||||||||| |||||| |||  ||||||||||||||| |||||||||
Db          3 ATATAAACCCAGGCATTCGAGCTGGCAACAGCAGCCCCCCTTTGGGTCCCTTCCCTTTGT 62

Qy        569 ATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCTTGCAACTGAA 619
```

```
               |||||.|  |||||||||  |||||||||||||||||||||:||||||| |
Db          63 ATGGGAG--CTGTTTTCATGCTATTTCACTCTATTAAATCTKGCAACTGCA 111

RESULT 19
US-08-979-847B-105
; Sequence 105, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;    INFORMATION FOR SEQ ID NO: 105:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1481 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 105:
US-08-979-847B-105

  Query Match             10.8%;  Score 68.4;  DB 4;  Length 1481;
  Best Local Similarity   87.2%;  Pred. No. 2.2e-12;
  Matches   75;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy           1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
               ||||||||||||||||||||||||||||||||||||||||||||  || ||||||| |||
Db        1396 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1455

Qy          61 CAAATTGTTCTTCAAATGGAGCACCA 86
               || || ||  | |||||||| | |||
Db        1456 CAGATGGTCTTACAAATGGAACCCCA 1481

RESULT 20
US-09-513-999C-13878
; Sequence 13878, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
```

```
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 13878
;   LENGTH: 342
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 88
;   OTHER INFORMATION: y=c or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 89
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 92
;   OTHER INFORMATION: y=c or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 103
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 109
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 277
;   OTHER INFORMATION: y=c or t
US-09-513-999C-13878

  Query Match              6.5%;  Score 41.4;  DB 4;  Length 342;
  Best Local Similarity   62.4%;  Pred. No. 0.00095;
  Matches   63;  Conservative    1;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          40 AGAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATG 99
               || |||||| |  | || ||||||||  |  | |||||||| |  || |||   ||
Db         190 AGGATCAAAGCCATCAAGCTACAAATGATCTTACAAATGGAACCTCAAATGAGCTCAGCT 249

Qy         100 ACTAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAGCCC 140
                      || |||| ||||||||||| :  || ||| | ||
Db         250 CACGGCTTCTACCGAGGACCCCTGGATYAACCCGCTGGTCC 290

RESULT 21
US-10-140-002-541
; Sequence 541, Application US/10140002
; Patent No. 6725730
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C59
;  CURRENT APPLICATION NUMBER: US/10/140,002
;  CURRENT FILING DATE:  2002-05-06
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 541
;   LENGTH: 3824
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-140-002-541

  Query Match             6.5%;  Score 41.2;  DB 4;  Length 3824;
  Best Local Similarity   61.8%;  Pred. No. 0.0052;
  Matches   84;  Conservative    0;  Mismatches   48;  Indels    4;  Gaps    1;

Qy        489 GCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCC 548
              | ||| ||| | ||    | ||   ||| |  |      ||||| |||  | ||||||||
Db       3693 GTGGGTGGGGCGAGACAAGAGAATAAAAGCAGGCTGCCTGAGCCAGCAGTGACAACCCCC 3752

Qy        549 TTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATC 608
               | ||||||||||||      |||  ||| |||| |  | || |||      | ||||||
Db       3753 CTCGGGTCCCCTCCCACGCCGTGGAAGCTTTGTTCTTTCGCTCTTTGC----AATAAATC 3808

Qy        609 TTGCAACTGAAAAAAA 624
              |||| ||||   ||||
Db       3809 TTGCTACTGCCCAAAA 3824


RESULT 22
US-09-120-653D-6
; Sequence 6, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
;  APPLICANT: YOON, JI-WON
;  APPLICANT:  JUN, HEE-SOOK
;  APPLICANT:  PARK, HAE-JOON
;  APPLICANT:  AHN, JONG-SEONG
;  APPLICANT:  HA, YOUNG-JU
;  APPLICANT:  CHUNG, SOO-IL
;  TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
;  FILE REFERENCE: 98-338
;  CURRENT APPLICATION NUMBER: US/09/120,653D
;  CURRENT FILING DATE:  1998-07-22
;  PRIOR APPLICATION NUMBER: KR 98-10108
;  PRIOR FILING DATE: 1998-03-24
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 6
;   LENGTH: 464
;   TYPE: DNA
;   ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-6

  Query Match             6.4%;  Score 40.4;  DB 3;  Length 464;
  Best Local Similarity   62.7%;  Pred. No. 0.0025;
  Matches   79;  Conservative    0;  Mismatches   46;  Indels    1;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||| | || |||||| | ||||| || ||||| |  || ||| || |  |  | || |||
Db        337 CCCAGCATTTTTAACGTTCTTGTCAAATTTGT-TTGTCTAGATTCGAGGCCTTCAAGCTA 395

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || ||  | || ||  | | ||  ||    || |  |   |  || || | |||| |
Db        396 CAGATGGTCTTACATATCAAACCCCGAATAATTTCAACTAACTACTTCTACGGAGGACAC 455

Qy        121 CTGGAC 126
              ||||||
```

```
Db        456 CTGGAC 461


RESULT 23
US-09-422-978-3579/c
; Sequence 3579, Application US/09422978
; Patent No. 6537751
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Chumakov, Ilya
;  TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
;  FILE REFERENCE: GENSET.020CP1
;  CURRENT APPLICATION NUMBER: US/09/422,978
;  CURRENT FILING DATE: 1999-10-20
;  EARLIER APPLICATION NUMBER: US 09/298,850
;  EARLIER FILING DATE: 1999-04-21
;  EARLIER APPLICATION NUMBER: US 60/109,732
;  EARLIER FILING DATE: 1998-11-23
;  EARLIER APPLICATION NUMBER: US 60/082,614
;  EARLIER FILING DATE: 1998-04-21
;  NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 3579
;   LENGTH: 47
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 24
;   OTHER INFORMATION: 99-6549-275 : polymorphic base A or G
US-09-422-978-3579

  Query Match             6.1%;  Score 38.6;  DB 4;  Length 47;
  Best Local Similarity   87.2%;  Pred. No. 0.0023;
  Matches   41;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        355 GTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACAC 401
              ||||| ||| |||||| ||||||:|||||||| |||| |||||||||
Db         47 GTGACCGCACCCACCTTTAAACAYGGGGCTTGTAACTCAGCTCACAC 1


RESULT 24
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             5.9%;  Score 37.6;  DB 1;  Length 7218;
  Best Local Similarity   3.7%;  Pred. No. 0.12;
  Matches   10;  Conservative  154;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        281 GAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAA 340
              :::::::::::  :::::: ::::  ::  :::      :::  :: ::::::     ::
Db       1322 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1263

Qy        341 GCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACA 400
              :   ::  ::::::: ::  : :   :     ::: : ::::   : ::   ::   : :
Db       1262 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1203

Qy        401 CCCGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAAT 460
                 ::  ::  ::::::   :  :::: :  ::  ::: ::::: :::::: :::::::
Db       1202 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1143

Qy        461 AGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCA 520
              ::  ::  :   :  :   ::::: : :: :::::::: :::::  :::: : :::   :
Db       1142 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1083

Qy        521 GGCATTCGAGCCGGCAACGGCAACCCCCTTTG 552
              :: :   :::  :: :|  |||| | || | |
Db       1082 RRRRRRRRRRRRRRRRRATCGCAAGCTCCCTCG 1051

RESULT 25
US-09-854-133-539
; Sequence 539, Application US/09854133
; Patent No. 6759508
; GENERAL INFORMATION:
;  APPLICANT: Lodes, Michael J.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.475C10
;  CURRENT APPLICATION NUMBER: US/09/854,133
;  CURRENT FILING DATE:  2001-05-11
;  NUMBER OF SEQ ID NOS: 735
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 539
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-854-133-539

  Query Match             5.7%;  Score 36;  DB 4;  Length 396;
  Best Local Similarity   50.6%;  Pred. No. 0.067;
  Matches   87;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy        132 TGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAAC 191
              | ||||  |||   || ||||       | |||   || |||||  | | | ||   |||
Db        224 TCCTAGAGCATAAACCCATGTGTGGCCAAGTGAGATCAGCCCTCAAGGGCACATGCCAAG 283
```

```
Qy        192 TGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAAC 251
               ||| | |  |   | |   |   ||| | ||| |   |   |   ||   || |   |
Db        284 GGCAGAGCAGCCCATGTAGACAGCTTCGGAGGGCATGGGGGTGTAGGGAGTTCGGGGTAG 343

Qy        252 CTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGG 303
              |||| ||  | |  || | |      ||  || || | | || || | ||||
Db        344 CTCCTCATTAACTATTTGTTGGGTGAGTAAAGGGGTGAGGCTCAGTGGCAGG 395

RESULT 26
US-09-220-132-94
; Sequence 94, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
;  APPLICANT:  Shyjan, Andrew W.
;  TITLE OF INVENTION:  METHODS AND COMPOSITONS FOR THE IDENTIFICATION AND ASSESSMENT
;  TITLE OF INVENTION:  OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCE
;  FILE REFERENCE:  07334-074001
;  CURRENT APPLICATION NUMBER: US/09/220,132
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  US 60/079,303
;  PRIOR FILING DATE:  1998-03-25
;  PRIOR APPLICATION NUMBER:  US 60/068,821
;  PRIOR FILING DATE:  1997-12-24
;  NUMBER OF SEQ ID NOS:  191
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 94
;   LENGTH: 1497
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1497)
;   OTHER INFORMATION: n = A,T,C or G
US-09-220-132-94

  Query Match             5.5%;  Score 35.2;  DB 4;  Length 1497;
  Best Local Similarity   19.9%;  Pred. No. 0.29;
  Matches   49;  Conservative   87;  Mismatches  110;  Indels    0;  Gaps    0;

Qy        233 AGAGCGGTCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGAC 292
              | :::: : :: | :| :: |   ::| :  ::||   |||||: : ::| :|||: : :
Db        268 AMMKMRTWKWKAAMMCCMMWTWWKSMAAMYWMYTTTTTTTTTCMGSMWRAARGGGRARSY 327

Qy        293 TGAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGA 352
              ::: :::    | :  :::| |||:: |    :|:   :||     :: :| :::
Db        328 KRMMWKRWYMAAAWYCMWKGKTTTYMAAAWKGRAMYKGRAAARRGSMWTYCCRSMMMMWY 387

Qy        353 AGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCA 412
                 : ::   :|  |  | ::|:  :  :|  :  :::    |:     ::  :::| ::
Db        388 TWWWAMMMMAWTTWAWTTTYMARKTYCMRGGSYTKRMMRGSCCYKGGGGSMWTYYWARSM 447

Qy        413 GAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCT 472
               :|:   :: :||: :| ::|::  |   :|| ::  ||:::        ::|:|  |
Db        448 ARGRWTGSMWWAAWTKGGYMAWYYTGKKGRAARRKTKGGKWMYTTTKGYTYMAWTTTTWA 507

Qy        473 ATTGCC 478
              :|: ||
Db        508 WTWMCC 513

RESULT 27
US-09-218-363-3/c
; Sequence 3, Application US/09218363
; Patent No. 6387616
; GENERAL INFORMATION:
;  APPLICANT: Ozelius, Laurie J.
;  APPLICANT: Breakefield, Xandra O.
;  TITLE OF INVENTION: TORSIN, TORSIN GENES, AND METHODS OF USE
;  FILE REFERENCE: MGH-1184pA2
;  CURRENT APPLICATION NUMBER: US/09/218,363
;  CURRENT FILING DATE: 1998-12-22
;  EARLIER APPLICATION NUMBER: 09/099,454
;  EARLIER FILING DATE: 1998-06-18
;  EARLIER APPLICATION NUMBER: 60/050,244
;  EARLIER FILING DATE: 1997-06-19
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 3568
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (994)...(1863)
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(3568)
;   OTHER INFORMATION: n = A,T,C or G
US-09-218-363-3

  Query Match             5.4%;  Score 34;  DB 3;  Length 3568;
  Best Local Similarity   53.6%;  Pred. No. 1.3;
  Matches   67;  Conservative    1;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        511 TATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTAT 570
              | ||||   |    |||| :||||   |||   | | | |    ||||    |||||
Db        650 TTTAAAAAWATTTTTTCGGSCCGGGCNCGGTGTCACGCCTGTAATCCCAGGACTTTGGGA 591

Qy        571 GGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAA 630
              || ||  | |   |||| | ||||||||| |          |||  |  |||||| |
Db        590 GGCCGAGCCGAGATCACGCCATTTCACTCCAGCCTGGGCAACAAGAGTGAAAAAATATCT 531

Qy        631 AAAAA 635
              |||||
Db        530 AAAAA 526


RESULT 28
US-09-621-976-9850
; Sequence 9850, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 9850
;   LENGTH: 117
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-9850

  Query Match             5.3%;  Score 33.8;  DB 4;  Length 117;
  Best Local Similarity   60.2%;  Pred. No. 0.17;
  Matches   56;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        543 ACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTAT 602
              | |||  |  |||| ||| || |  | |||  ||| |||| |  | || |||      |
Db          3 AACCCTCTCAGGTCACCTTCCATGCTGTGGTGGCTTTGTTCTTTCCCTCTTTGCAATAAA 62

Qy        603 TAAATCTTGCAACTGAAAAAAAAAAAAAAAAAAA 635
              ||   ||        ||||||||||||||||||
Db         63 TATTGCTGCTGTTCAAAAAAAAAAAAAAAAAAAA 95


RESULT 29
US-09-621-976-1047
; Sequence 1047, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
```

```
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 1047
;   LENGTH: 177
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 24..176
US-09-621-976-1047

  Query Match             5.3%;  Score 33.8;  DB 4;  Length 177;
  Best Local Similarity   60.2%;  Pred. No. 0.22;
  Matches   56;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        543 ACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTAT 602
              | |||  |  |||| ||| || |  | |||  ||| |||| |  | || |||      |
Db          3 AACCCTCTCAGGTCACCTTCCATGCTGTGGTGGCTTTGTTCTTTCCCTCTTTGCAATAAA 62

Qy        603 TAAATCTTGCAACTGAAAAAAAAAAAAAAAAAAA 635
              ||   ||        ||||||||||||||||||
Db         63 TATTGCTGCTGTTCAAAAAAAAAAAAAAAAAAAA 95


RESULT 30
US-08-011-398B-1/c
; Sequence 1, Application US/08011398B
; Patent No. 5512473
;  GENERAL INFORMATION:
;    APPLICANT:  Roger Brent
;    APPLICANT:  Antonis S. Zervos
;    TITLE OF INVENTION:  MAX-INTERACTING PROTEINS AND RELATED
;    TITLE OF INVENTION:  MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44  Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/011,398B
;      FILING DATE:  29 JAN 1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul T. Clark
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/160001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2417
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
US-08-011-398B-1

  Query Match             5.2%;  Score 33;  DB 1;  Length 2417;
  Best Local Similarity   58.8%;  Pred. No. 2.1;
  Matches   57;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        539 GGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACT 598
              || ||  ||  ||||   || | ||  ||  ||   |  || | |  ||| ||  | | |
Db        162 GGGAAGGCCTCTTGGAGACCTTACCCCTGGCTGTTTGGACTTTGTATACTTTAAATAATT 103

Qy        599 CTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
                | ||    |    ||| ||||||||||||||||||
Db        102 TAACTACCCTTAATTACTTAAAAAAAAAAAAAAAAAA 66

Search completed: January 22, 2005, 19:36:55
Job time : 61.7997 secs
```

```
                        GenCore version 5.1.6
               Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 12:12:40 ; Search time 344.872 Seconds
                                           (without alignments)
                                           10579.682 Million cell updates/sec

Title:          US-09-319-156B-6
Perfect score:  635
Sequence:       1 ccctgtatctttaacctcct..........tgaaaaaaaaaaaaaaaaa 635

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters:      8600550

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
                14:  /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
                16:  /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
                17:  /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
                18:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                19:  /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
                20:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                21:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                         Description
-------------------------------------------------------------------------
    1     635   100.0    635  8  US-08-979-847-102          Sequence 102, App
    2     635   100.0    635 15  US-10-114-104-102          Sequence 102, App
    3   617.4    97.2   2030 16  US-10-637-565-18           Sequence 18, Appl
c   4   564.8    88.9 366710 18  US-10-719-993-6792         Sequence 6792, Ap
    5   546.4    86.0   2074 16  US-10-416-642-4            Sequence 4, Appli
    6     545    85.8   1329  8  US-08-979-847-108          Sequence 108, App
    7     545    85.8   1329 15  US-10-114-104-108          Sequence 108, App
    8   543.4    85.6  21646 13  US-10-087-192-910          Sequence 910, App
    9   533.8    84.1   2946 13  US-10-114-893-134          Sequence 134, App
   10   533.8    84.1   2946 15  US-10-016-249-3            Sequence 3, Appli
c  11   532.8    83.9 326014  9  US-09-731-231A-3           Sequence 3, Appli
c  12   532.8    83.9 326014 17  US-10-751-985-3            Sequence 3, Appli
```

```
   13   525.8    82.8   1136 16  US-10-632-793-25           Sequence 25, Appl
   14   522.8    82.3 283351 18  US-10-719-993-7065         Sequence 7065, Ap
   15   522.8    82.3 1980090 18  US-10-719-993-6815        Sequence 6815, Ap
   16   522.6    82.3   2782 16  US-10-632-793-26           Sequence 26, Appl
   17     521    82.0   2782 16  US-10-133-036-1            Sequence 1, Appli
   18   517.4    81.5   2930  9  US-09-902-535-1            Sequence 1, Appli
   19   516.2    81.3  56093 10  US-09-873-367C-81          Sequence 81, Appl
   20   511.4    80.5   7582 16  US-10-632-793-30           Sequence 30, Appl
c  21   510.6    80.4 161334 13  US-10-087-192-730          Sequence 730, App
   22   500.8    78.9 285020 13  US-10-087-192-1666         Sequence 1666, Ap
   23   499.6    78.7   1894  9  US-09-864-761-4444         Sequence 4444, Ap
   24   495.6    78.0   3372 16  US-10-632-793-28           Sequence 28, Appl
   25   493.2    77.7   2052 16  US-10-276-774-678          Sequence 678, App
   26   482.6    76.0  22436 11  US-09-997-722-148          Sequence 148, App
   27   481.4    75.8    849 16  US-10-220-120-15           Sequence 15, Appl
   28     466    73.4   2349 15  US-10-104-047-1148         Sequence 1148, Ap
   29     449    70.7   1564 15  US-10-204-887-10           Sequence 10, Appl
   30   446.2    70.3   1393 16  US-10-220-120-17           Sequence 17, Appl
   31   434.4    68.4   7974 14  US-10-198-846-9936         Sequence 9936, Ap
   32   427.6    67.3  44063 17  US-10-322-281-718          Sequence 718, App
   33   424.6    66.9    604 13  US-10-027-632-289680       Sequence 289680,
   34   424.6    66.9    604 15  US-10-027-632-289680       Sequence 289680,
   35   415.6    65.4   1394 17  US-10-466-531-47           Sequence 47, Appl
c  36   395.8    62.3    548 15  US-10-029-386-8554         Sequence 8554, Ap
c  37   388.4    61.2  23855 15  US-10-242-355-1175         Sequence 1175, Ap
c  38   383.4    60.4    657 13  US-10-027-632-224966       Sequence 224966,
c  39   383.4    60.4    657 15  US-10-027-632-224966       Sequence 224966,
c  40     373    58.7    439  9  US-09-764-881-183          Sequence 183, App
c  41     373    58.7    439 10  US-09-764-881-183          Sequence 183, App
c  42     373    58.7    439 16  US-10-242-747-183          Sequence 183, App
c  43     373    58.7    439 16  US-10-158-057-365          Sequence 365, App
c  44     371    58.4    587 13  US-10-027-632-321647       Sequence 321647,
c  45     371    58.4    587 15  US-10-027-632-321647       Sequence 321647,


                              ALIGNMENTS


RESULT 1
US-08-979-847-102
; Sequence 102, Application US/08979847
; Publication No. US20030039664A1
;  GENERAL INFORMATION:
;    APPLICANT:  PERRON, HERVE
;    APPLICANT:  BESEME, FREDERIC
;    APPLICANT:  BEDIN, FREDERIC
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  KOMURIAN-PRADEL, FLORENCE
;    APPLICANT:  JOLIVET-REYNAUD, COLETTE
;    APPLICANT:  MANDRAND, BERNARD
;    APPLICANT:  GARSON, JEREMY
;    APPLICANT:  TUKE, PHILIP
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  210
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/979,847
;      FILING DATE:  26-NOV-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  BERRIDGE, WILLIAM P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 39046A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  102:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  635 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-979-847-102

  Query Match             100.0%;  Score 635;  DB 8;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-187;
  Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||||||||||||||
Db        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635


RESULT 2
US-10-114-104-102
; Sequence 102, Application US/10114104
; Publication No. US20030198647A1
;   GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
```

```
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/114,104
;             FILING DATE: 03-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/979,847
;             FILING DATE: 26-NOV-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;    INFORMATION FOR SEQ ID NO: 102:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 635 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-10-114-104-102

  Query Match             100.0%;  Score 635;  DB 15;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-187;
  Matches  635;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
```

```
Qy         361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy         421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy         481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy         541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy         601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
               |||||||||||||||||||||||||||||||||||
Db         601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
```

RESULT 3
US-10-637-565-18
; Sequence 18, Application US/10637565
; Publication No. US20040043381A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT: PERRON, Herve
;  APPLICANT: KOMURIAN-PRADEL, Florence
;  TITLE OF INVENTION: THE LTR REGION OF MSRV-1 AND THE PROTEINS IT ENCODES, AND PROBES
;  TITLE OF INVENTION: METHODS FOR DETECTING MSRV-1 RETROVIRUS
;  FILE REFERENCE: 110257
;  CURRENT APPLICATION NUMBER: US/10/637,565
;  CURRENT FILING DATE: 2003-08-11
;  PRIOR APPLICATION NUMBER: US/09/890,340
;  PRIOR FILING DATE: 2001-11-27
;  PRIOR APPLICATION NUMBER: PCT/IB00/00159
;  PRIOR FILING DATE: 2000-02-15
;  PRIOR APPLICATION NUMBER: EP 99420041.8
;  PRIOR FILING DATE: 1999-02-15
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 2030
;   TYPE: DNA
;   ORGANISM: MSRV-1 retrovirus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1626)
US-10-637-565-18

```
  Query Match             97.2%;  Score 617.4;  DB 16;  Length 2030;
  Best Local Similarity   98.3%;  Pred. No. 8.3e-182;
  Matches  624;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy           1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
               ||||||||||||||||||||||||||||||||||||||||||||  || ||||||| |||
Db        1396 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1455

Qy          61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
               || || ||  | |||||||| | |||||||||||||||||||||||||||||||||||||
Db        1456 CAGATGGTCTTACAAATGGAACCCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 1515

Qy         121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1516 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 1575

Qy         181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1576 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 1635

Qy         241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1636 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 1695

Qy         301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1696 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 1755

Qy         361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1756 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 1815

Qy         421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1816 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 1875

Qy         481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1876 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 1935

Qy         541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1936 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 1995

Qy         601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
               |||||||||||||||||||||||||||||||||||
Db        1996 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 2030
```

RESULT 4
US-10-719-993-6792/c
; Sequence 6792, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE: 2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6792
;   LENGTH: 366710
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(366710)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6792

```
  Query Match             88.9%;  Score 564.8;  DB 18;  Length 366710;
  Best Local Similarity   93.4%;  Pred. No. 2.6e-164;
  Matches  590;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy           1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
               |||||||||||||||| ||||||||||||||| ||||||||||| |||||||||||||||
Db       43899 CCCTGTATCTTTAACCACCTTGTTAAGTTTGTGTCTTCCAGAATTAAAACTGTAAAACTA 43840

Qy          61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
               ||||| |||||||||||||| | ||||||| |||||||| |||||||| | |  ||||||
Db       43839 CAAATCGTTCTTCAAATGGAACCCCAGATGCAGTCCATGTCTAAGATCTCCCACGGACCC 43780

Qy         121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
               || ||| ||||||||||||||||||| | |||||||||||| |||| ||| ||||| |||
Db       43779 CTAGACTGGCCTGCTAGCCCATGCTCTGGTGTTAATGACATCGAAGTCACTCCTCCTGAG 43720

Qy         181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
               |||||||||||||||||||||||||||||||||| |||||| |||||||||||||| |||
Db       43719 GAAATCTCAACTGCACAACCCCTACTATGCCCCAGTTCAGCAGGAAGCAGTTAGAGTGGT 43660

Qy         241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
               |||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db       43659 CATCAGCCAACCTCACCAACAGCACTTGGGTTTTCCTGTTGAGAGAGGGGACTGAGAGAC 43600
```

```
Qy           301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
                 |||||||||||||||||||||||| |  |||||||||||| ||||||||||||||||||
Db         43599 AGGACTAGCTGGATTTCCTAGGCCGATTAAGAATCCCTAAACCTAGCTGGGAAGGTGACC 43540

Qy           361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                 |||||||||| |||||| ||||||||||||||||||||||||| ||||||||||||||||
Db         43539 GCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAACCAATCAGAGAGCTC 43480

Qy           421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
                 |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db         43479 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 43420

Qy           481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
                 ||||||||| ||||||||||||||| |  ||||||||||||||||||||||| |||||||
Db         43419 AGAGCACAGTGGGAGGGACAAGGATTGCAATATAAACCCAGGCATTCGAGCCAGCAACGG 43360

Qy           541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
                 ||||| |||||||||||| |||||||||||||| ||||||||||||||||||||||||||
Db         43359 CAACCGCCTTTGGGTCCCTTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 43300

Qy           601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAA 632
                 ||||||||||||||||||||||||||||||||
Db         43299 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAA 43268

RESULT 5
US-10-416-642-4
; Sequence 4, Application US/10416642
; Publication No. US20040043452A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  RAMKUMAR, Jayalaxmi
;  APPLICANT:  ARVIZU, Chandra
;  TITLE OF INVENTION: EMBRYOGENESIS ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-0842 PCT
;  CURRENT APPLICATION NUMBER: US/10/416,642
;  CURRENT FILING DATE:  2003-05-13
;  PRIOR APPLICATION NUMBER: 60/249,407
;  PRIOR FILING DATE: 2000-11-15
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PERL Program
; SEQ ID NO 4
;   LENGTH: 2074
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20040043452A1 7477736CB1
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1995
;   OTHER INFORMATION: a, t, c, g, or other
US-10-416-642-4

  Query Match             86.0%;  Score 546.4;  DB 16;  Length 2074;
  Best Local Similarity   93.8%;  Pred. No. 1.3e-159;
  Matches  579;  Conservative    0;  Mismatches   37;  Indels    1;  Gaps    1;

Qy             1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
                 ||||||||||||||||||||||||||||||||||||||||||||  || ||||||| |||
Db          1459 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1518

Qy            61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                 ||||| |||||||||||||||| ||||||| ||||||||||||| ||| |||  ||||||
Db          1519 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAAATCTACCACGGACCC 1578

Qy           121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
                 ||||||||||||||||||||||||||||||||||||||||| |||||||| |||||| ||
Db          1579 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACTCCTCCCAAG 1638

Qy           181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
                 ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          1639 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 1698

Qy           241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                 | |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1699 CGTCAGTCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 1758

Qy           301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
                 |||||||||||||||||||||||| |  |||||||||||||||||||||||||||||||
Db          1759 AGGACTAGCTGGATTTCCTAGGCCGATTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 1818

Qy           361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                 || ||||||| |||||| ||||||||||||||||||||||||| ||||||||||||||||
Db          1819 GCGTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAACCAATCAGAGAGCTC 1878

Qy           421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
                 |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1879 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 1938

Qy           481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
                 ||||||||| ||||||||||||||| |  ||||||||||||||||||||||| || |  |
Db          1939 AGAGCACAGTGGGAGGGACAAGGATTGCAATATAAACCCAGGCATTCGAGCCAGC-ANAG 1997

Qy           541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
                 ||||| |||||||||||| |||| ||||||||| ||||||||||||||||||||||||||
Db          1998 CAACCGCCTTTGGGTCCCTTCCCCTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 2057

Qy           601 ATTAAATCTTGCAACTG 617
                 |||||||||||||||||
Db          2058 ATTAAATCTTGCAACTG 2074

RESULT 6
US-08-979-847-108
; Sequence 108, Application US/08979847
; Publication No. US20030039664A1
;  GENERAL INFORMATION:
;    APPLICANT:  PERRON, HERVE
;    APPLICANT:  BESEME, FREDERIC
;    APPLICANT:  BEDIN, FREDERIC
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  KOMURIAN-PRADEL, FLORENCE
;    APPLICANT:  JOLIVET-REYNAUD, COLETTE
;    APPLICANT:  MANDRAND, BERNARD
;    APPLICANT:  GARSON, JEREMY
;    APPLICANT:  TUKE, PHILIP
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTI
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  210
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/979,847
;      FILING DATE:  26-NOV-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERRIDGE, WILLIAM P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 39046A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  108:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1329 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-979-847-108

  Query Match             85.8%;  Score 545;  DB 8;  Length 1329;
  Best Local Similarity   92.4%;  Pred. No. 2.8e-159;
  Matches  572;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||| ||| ||||||||||||||||||||||||||||  || ||||||| |||
Db        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||| | |||||| ||||||||||||||||||| |||||||||||||
Db        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||| ||||||| |||||||||||||
Db        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||| ||||||||| ||||||||||| |||||||||||||||||||||||
Db        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||||||| |||||||||| ||||||||||| ||||||||| |||| ||||| |
Db        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856

Qy        601 ATTAAATCTTGCAACTGAA 619
              |||||||| |||||||| |
Db        857 ATTAAATCATGCAACTGCA 875


RESULT 7
US-10-114-104-108
; Sequence 108, Application US/10114104
; Publication No. US20030198647A1
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
```

```
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/114,104
;            FILING DATE: 03-Apr-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/979,847
;            FILING DATE: 26-NOV-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;  INFORMATION FOR SEQ ID NO: 108:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1329 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 108:
US-10-114-104-108

  Query Match             85.8%;  Score 545;  DB 15;  Length 1329;
  Best Local Similarity   92.4%;  Pred. No. 2.8e-159;
  Matches  572;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||| ||| ||||||||||||||||||||||||||||  || ||||||| |||
Db        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||| | |||||| ||||||||||||||||||| |||||||||||||
Db        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||| ||||||| |||||||||||||
Db        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||| ||||||||| ||||||||||| |||||||||||||||||||||||
Db        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736
```

```
Qy         481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
               |||||||||||||| |||||||||| ||||||||||| ||||||||| |||| ||||| |
Db         737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796

Qy         541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
               ||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db         797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856

Qy         601 ATTAAATCTTGCAACTGAA 619
               |||||||| |||||||| |
Db         857 ATTAAATCATGCAACTGCA 875


RESULT 8
US-10-087-192-910
; Sequence 910, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 910
;   LENGTH: 21646
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(21646)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-910

  Query Match             85.6%;  Score 543.4;  DB 13;  Length 21646;
  Best Local Similarity   94.8%;  Pred. No. 3.3e-158;
  Matches  562;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy           1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
               ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db       19571 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAACTGTAAAACTA 19630

Qy          61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
               ||||| |||||||||||||||| ||||||| |||||||||||||||||||||   |||||
Db       19631 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACCACAGACCC 19690

Qy         121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
               ||||||| ||||||||||||||||||||||||||||||||| ||||||||||| ||||||
Db       19691 CTGGACCAGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCCCCCGAG 19750

Qy         181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
               |||||||||||||  |||||||||||||||||||||||||| ||||||||||| ||||||
Db       19751 GAAATCTCAACTGTGCAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTACAGCGGT 19810

Qy         241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       19811 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGCACTGAGAGAC 19870

Qy         301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
               | |||||||||||||||||||||| || |||||||||||||| |||||||||||||||||
Db       19871 AAGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCGTAGCTGGGAAGGTGACT 19930

Qy         361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               |||||||||| |||||||||||||||| ||||||||||||||| ||||||||||||||||
Db       19931 GCATCCACCTTTAAACATGGGGCTTGCCACTTAGCTCACACCCAACCAATCAGAGAGCTC 19990
```

```
Qy         421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
               |||||||||||||||||||||||| |||||||||||||||||||||||| || |||||||
Db       19991 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCACCTTTTGCCTG 20050

Qy         481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
               |||||||||| |||||||||||||||||||||||||||| ||||||| ||||||||| ||
Db       20051 AGAGCACAGCAGGAGGGACAAGGATCGGGATATAAACCCGGGCATTCAAGCCGGCAATGG 20110

Qy         541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATT 593
               ||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       20111 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATT 20163


RESULT 9
US-10-114-893-134
; Sequence 134, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT: McCoy, John M.
;  APPLICANT: LaVallie, Edward R.
;  APPLICANT: Collins-Racie, Lisa A.
;  APPLICANT: Evans, Cheryl
;  APPLICANT: Merberg, David
;  APPLICANT: Treacy, Maurice
;  APPLICANT: Bowman, Michael R.
;  APPLICANT: Spaulding, Vikki
;  APPLICANT: Carlin-Duckett, McKeough
;  APPLICANT: Kelleher, Kerry S.
;  APPLICANT: Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: GI 6000-10A
;  CURRENT APPLICATION NUMBER: US/10/114,893
;  CURRENT FILING DATE: 2002-04-02
;  EARLIER APPLICATION NUMBER: 09/413,232
;  EARLIER FILING DATE: 1999-10-06
;  NUMBER OF SEQ ID NOS: 321
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 134
;   LENGTH: 2946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-114-893-134

  Query Match             84.1%;  Score 533.8;  DB 13;  Length 2946;
  Best Local Similarity   91.5%;  Pred. No. 1.3e-155;
  Matches  581;  Conservative    0;  Mismatches   42;  Indels   12;  Gaps    1;

Qy           1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
               |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        2323 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2382

Qy          61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                           |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        2383 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2430

Qy         121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
               ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db        2431 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2490

Qy         181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
               ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        2491 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2550

Qy         241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
               | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db        2551 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2610

Qy         301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
               |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db        2611 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2670
```

```
Qy          361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                 ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db         2671 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2730

Qy          421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
                |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db         2731 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2790

Qy          481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
                |||||||||| |||||||||| |||||||||||||||||| |   ||||||||||||||||
Db         2791 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2850

Qy          541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
                ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db         2851 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2910

Qy          601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
                ||||||||||||||||| |||||||||||||||||
Db         2911 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2945


RESULT 10
US-10-016-249-3
; Sequence 3, Application US/10016249
; Publication No. US20030100053A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Mi, Sha
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6006B.AJ172A
;  CURRENT APPLICATION NUMBER: US/10/016,249
;  CURRENT FILING DATE:  2001-10-30
;  PRIOR APPLICATION NUMBER: US/09/175,928
;  PRIOR FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-016-249-3

  Query Match             84.1%;  Score 533.8;  DB 15;  Length 2946;
  Best Local Similarity   91.5%;  Pred. No. 1.3e-155;
  Matches  581;  Conservative    0;  Mismatches   42;  Indels   12;  Gaps    1;

Qy            1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
                |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db         2323 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2382

Qy           61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                            |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db         2383 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2430

Qy          121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
                ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db         2431 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2490

Qy          181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
                ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db         2491 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2550

Qy          241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db         2551 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2610
```

```
Qy          301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
                |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||| 
Db         2611 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2670

Qy          361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                 ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db         2671 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2730

Qy          421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
                |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db         2731 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2790

Qy          481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
                |||||||||| |||||||||| |||||||||||||||||| |   ||||||||||||||||
Db         2791 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2850

Qy          541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
                ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db         2851 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2910

Qy          601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
                ||||||||||||||||| |||||||||||||||||
Db         2911 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2945


RESULT 11
US-09-731-231A-3/c
; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1
; GENERAL INFORMATION:
;  APPLICANT: GUEGLER, Karl et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001007
;  CURRENT APPLICATION NUMBER: US/09/731,231A
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 326014
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(326014)
;   OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3

  Query Match             83.9%;  Score 532.8;  DB 9;  Length 326014;
  Best Local Similarity   90.5%;  Pred. No. 2.4e-154;
  Matches  583;  Conservative    0;  Mismatches   52;  Indels    9;  Gaps    1;

Qy            1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
                |||||||||||||||||||||||||||||||||||||||| |||  || || ||||||||
Db       170822 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAAAATTGAAGCTCTAAAACTA 170763

Qy           61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                |||||||||||||||||||||| ||||| | ||||||||||||||||| ||||  |||||
Db       170762 CAAATTGTTCTTCAAATGGAGCCCCAGACGCAGTCCATGACTAAGATCTACCGCAGACCC 170703

Qy          121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
                ||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db       170702 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCTCCAGAG 170643

Qy          181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
                ||||||||||||||| |||||| |||| ||||||||||||| ||||||||||||||| ||
Db       170642 GAAATCTCAACTGCATAACCCCCACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 170583

Qy          241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |||| |||| ||||||||||||||||||||||||||||||||||||||| ||| ||||||
Db       170582 CATCGGCCATCCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTCAGAGAC 170523
```

```
Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||||
Db     170522 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 170463

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
              || || |||| |||||  |||||||||||||||||||||||| ||||||||
Db     170462 GCTTCTACCTTTAAACCCGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGGTAGGAA 170403

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              |||||||||||||||||| ||| ||||| |||| ||||||||||||||||||||||||||
Db     170402 AGAGAGCTCACTAAAATGGTAACTAGGCTAAAACAGGAGGTAAAGAAATAGCCAATCATC 170343

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              ||| |||||||| ||||| ||||||||||| ||| ||| |||||||||||||||| ||||
Db     170342 TATCGCCTGAGAACACAGTGGGAGGGACAATGATTGGGTTATAAACCCAGGCATTGGAGC 170283

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              | |||| ||| |||   ||||||||||||||||||||||||| |||||||||||||||||
Db     170282 CAGCAATGGCTACCATTTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTA 170223

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||| | ||||||||||||||| |
Db     170222 TTTCACTCTATTAAATCTTGCAACAGCAAAAAAAAAAAAAAAAGA 170179


RESULT 12
US-10-751-985-3/c
; Sequence 3, Application US/10751985
; Publication No. US20040126861A1
; GENERAL INFORMATION:
;  APPLICANT: GUEGLER, Karl et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001007CON
;  CURRENT APPLICATION NUMBER: US/10/751,985
;  CURRENT FILING DATE:  2004-01-07
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 326014
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(326014)
;   OTHER INFORMATION: n = A,T,C or G
US-10-751-985-3

  Query Match             83.9%;  Score 532.8;  DB 17;  Length 326014;
  Best Local Similarity   90.5%;  Pred. No. 2.4e-154;
  Matches  583;  Conservative    0;  Mismatches   52;  Indels    9;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||| |||  || || ||||||||
Db     170822 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAAAATTGAAGCTCTAAAACTA 170763

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              |||||||||||||||||||||| ||||| | ||||||||||||||||| ||||  |||||
Db     170762 CAAATTGTTCTTCAAATGGAGCCCCAGACGCAGTCCATGACTAAGATCTACCGCAGACCC 170703

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db     170702 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCTCCAGAG 170643

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||||||||| |||||| |||| ||||||||||||| ||||||||||||||| ||
Db     170642 GAAATCTCAACTGCATAACCCCCACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 170583

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||| |||| ||||||||||||||||||||||||||||||||||||||| ||| ||||||
Db     170582 CATCGGCCATCCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTCAGAGAC 170523
```

```
Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||||
Db     170522 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 170463

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
              || || |||| |||||  |||||||||||||||||||||||| ||||||||
Db     170462 GCTTCTACCTTTAAACCCGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGGTAGGAA 170403

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              |||||||||||||||||| ||| ||||| |||| ||||||||||||||||||||||||||
Db     170402 AGAGAGCTCACTAAAATGGTAACTAGGCTAAAACAGGAGGTAAAGAAATAGCCAATCATC 170343

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              ||| |||||||| ||||| ||||||||||| ||| ||| |||||||||||||||| ||||
Db     170342 TATCGCCTGAGAACACAGTGGGAGGGACAATGATTGGGTTATAAACCCAGGCATTGGAGC 170283

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              | |||| ||| |||   ||||||||||||||||||||||||| |||||||||||||||||
Db     170282 CAGCAATGGCTACCATTTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTA 170223

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||||||||||||||||||| | ||||||||||||||| |
Db     170222 TTTCACTCTATTAAATCTTGCAACAGCAAAAAAAAAAAAAAAAGA 170179


RESULT 13
US-10-632-793-25
; Sequence 25, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 1136
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-632-793-25

  Query Match             82.8%;  Score 525.8;  DB 16;  Length 1136;
  Best Local Similarity   90.7%;  Pred. No. 2.6e-153;
  Matches  576;  Conservative    0;  Mismatches   47;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        510 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 569

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        570 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 617

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db        618 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 677

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        678 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 737

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
```

```
               | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db         738 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 797

Qy         301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
               |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db         798 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 857

Qy         361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                ||||||||| |||||| |||||||||||||||| ||||||| |||||||||||||||||
Db         858 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGTTCACACCTGACCAATCAGAGAGCTC 917

Qy         421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
               |||||||||||||||||||||| | |||||||||||||||||||||||||||||||| ||
Db         918 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCATG 977

Qy         481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
               |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db         978 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 1037

Qy         541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
               ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db        1038 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 1097

Qy         601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
               ||||||||||||| |||  ||||||||||||||||
Db        1098 ATTAAATCTTGCAGCTGCGAAAAAAAAAAAAAAAA 1132

RESULT 14
US-10-719-993-7065
; Sequence 7065, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7065
;   LENGTH: 283351
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(283351)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-7065

  Query Match             82.3%;  Score 522.8;  DB 18;  Length 283351;
  Best Local Similarity   93.4%;  Pred. No. 3e-151;
  Matches  568;  Conservative    1;  Mismatches   33;  Indels    6;  Gaps    2;

Qy          27 GTTTGTCTCTTCCAGAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATGGAGCACCA 86
               ||||||||||||||||||  ||||||||||||||||||| |||||||||||||||| |||
Db      174805 GTTTGTCTCTTCCAGAATTGAAACTGTAAAACTACAAATGGTTCTTCAAATGGAGCCCCA 174864

Qy          87 GATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAGCCCATGCTC 146
               |||| ||||||||||||||||||||||  |||||||||||:|||||||||||||||||||
Db      174865 GATGCAGTCCATGACTAAGATCCACCGCAGACCCCTGGACYGGCCTGCTAGCCCATGCTC 174924

Qy         147 CGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACT 206
               | ||||||||||||| |||||||||| ||| ||||||||||||||| |||||| ||||||
Db      174925 CAATGTTAATGACATGGAAGGCACCCTTCCGGAGGAAATCTCAACTACACAACACCTACT 174984

Qy         207 ATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCACT 266
               ||||||||||||||| |||||| |||||||| |||||  |||||||||||||||||||||
Db      174985 ATGCCCCAATTCAGCAGGAAGCTGTTAGAGCAGTCATTGGCCAACCTCCCCAACAGCACT 175044

Qy         267 TGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAA 326
               ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |
```

```
Db      175045 TGGGTTTTCCTGTTGAGAGAGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGA 175104

Qy         327 CGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTG 386
               | ||||||||||||||||||||||||||||||| |||||||||| |||||| ||||||||
Db      175105 CTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCGCATCCACCTTTAAACACGGGGCTTG 175164

Qy         387 CAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAATA 446
               |||||||||||||||    |||| ||||||| ||||||||||||||||||||||||||||
Db      175165 CAACTTAGCTCACAC----CCAACCAGAGAGTTCACTAAAATGCTAATTAGGCAAAAATA 175220

Qy         447 GGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATC 506
               ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      175221 GGAGGTACAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATC 175280

Qy         507 GGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTT 566
                |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      175281 AGGATATAAACCCAGGCATTCGAGCTGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTT 175340

Qy         567 GTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCTTGCAACTGAAAAAAAAA 626
               ||||||| ||||||||||||  || |||||||||||||||||||||||||| |||||| |
Db      175341 GTATGGGAGCTCTGTTTTCA--CTGTTTCACTCTATTAAATCTTGCAACTGCAAAAAAGA 175398

Qy         627 AAAAAAAA 634
               ||||||||
Db      175399 AAAAAAAA 175406

RESULT 15
US-10-719-993-6815
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6815
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6815

  Query Match             82.3%;  Score 522.8;  DB 18;  Length 1980090;
  Best Local Similarity   93.4%;  Pred. No. 7.6e-151;
  Matches  568;  Conservative    1;  Mismatches   33;  Indels    6;  Gaps    2;

Qy          27 GTTTGTCTCTTCCAGAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATGGAGCACCA 86
               ||||||||||||||||||  ||||||||||||||||||| |||||||||||||||| |||
Db     1777083 GTTTGTCTCTTCCAGAATTGAAACTGTAAAACTACAAATGGTTCTTCAAATGGAGCCCCA 1777142

Qy          87 GATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAGCCCATGCTC 146
               |||| ||||||||||||||||||||||  |||||||||||:|||||||||||||||||||
Db     1777143 GATGCAGTCCATGACTAAGATCCACCGCAGACCCCTGGACYGGCCTGCTAGCCCATGCTC 1777202

Qy         147 CGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACT 206
               | ||||||||||||| |||||||||| ||| ||||||||||||||| |||||| ||||||
Db     1777203 CAATGTTAATGACATGGAAGGCACCCTTCCGGAGGAAATCTCAACTACACAACACCTACT 1777262

Qy         207 ATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCACT 266
               ||||||||||||||| |||||| |||||||| |||||  |||||||||||||||||||||
Db     1777263 ATGCCCCAATTCAGCAGGAAGCTGTTAGAGCAGTCATTGGCCAACCTCCCCAACAGCACT 1777322

Qy         267 TGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAA 326
               ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |
Db     1777323 TGGGTTTTCCTGTTGAGAGAGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGA 1777382
```

```
Qy        327 CGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTG 386
              | ||||||||||||||||||||||||||||||| |||||||||| |||||| ||||||||
Db    1777383 CTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCGCATCCACCTTTAAACACGGGGCTTG 1777442

Qy        387 CAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAATA 446
              |||||||||||||||    |||| ||||||| ||||||||||||||||||||||||||||
Db    1777443 CAACTTAGCTCACAC----CCAACCAGAGAGTTCACTAAAATGCTAATTAGGCAAAAATA 1777498

Qy        447 GGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATC 506
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1777499 GGAGGTACAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATC 1777558

Qy        507 GGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTT 566
               |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    1777559 AGGATATAAACCCAGGCATTCGAGCTGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTT 1777618

Qy        567 GTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCTTGCAACTGAAAAAAAAA 626
              ||||||| ||||||||||||  || |||||||||||||||||||||||||| |||||| |
Db    1777619 GTATGGGAGCTCTGTTTTCA--CTGTTTCACTCTATTAAATCTTGCAACTGCAAAAAAGA 1777676

Qy        627 AAAAAAAA 634
              ||||||||
Db    1777677 AAAAAAAA 1777684
```

```
RESULT 16
US-10-632-793-26
; Sequence 26, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
;   LENGTH: 2782
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-632-793-26

  Query Match             82.3%;  Score 522.6;  DB 16;  Length 2782;
  Best Local Similarity   90.4%;  Pred. No. 4e-152;
  Matches  574;  Conservative    0;  Mismatches   49;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||| ||||||||||||| |||||||||||||||||| || |||| ||||||
Db       2158 CCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTGAAACTA 2217

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2218 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2265

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2266 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2325

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2326 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2385
```

```
Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2386 GGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2445

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||| ||||||||| ||||||||||||
Db       2446 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCTAGGTGGGAAGGTGACC 2505

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2506 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2565

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | ||||||||||||||||||||||||| |||||||||
Db       2566 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATTTATTGCCTG 2625

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| ||||||||| |||||||||||||||||| |  |||||||||||||||||
Db       2626 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTCGAGCCGGCAACGG 2685

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       2686 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2745

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       2746 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2780
```

```
RESULT 17
US-10-133-036-1
; Sequence 1, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
;  APPLICANT: Conrad, Bernard
;  APPLICANT:  Mach, Bernard
;  TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
;  FILE REFERENCE: 23135-507
;  CURRENT APPLICATION NUMBER: US/10/133,036
;  CURRENT FILING DATE:  2002-04-26
;  PRIOR APPLICATION NUMBER: PCT/EP00/10659
;  PRIOR FILING DATE: 2000-10-30
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2782
;   TYPE: DNA
;   ORGANISM: Human endogenous retrovirus
US-10-133-036-1

  Query Match             82.0%;  Score 521;  DB 16;  Length 2782;
  Best Local Similarity   90.2%;  Pred. No. 1.3e-151;
  Matches  573;  Conservative    0;  Mismatches   50;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||| ||||||||||||| |||||||||||||||||| || |||| ||||||
Db       2158 CCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTGAAACTA 2217

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| || |  |||||
Db       2218 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACGGAAGACCC 2265

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2266 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2325

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2326 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2385

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2386 GGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2445
```

```
Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||| ||||||||| ||||||||||||
Db       2446 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCTAGGTGGGAAGGTGACC 2505

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2506 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2565

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | ||||||||||||||||||||||||| |||||||||
Db       2566 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATTTATTGCCTG 2625

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2626 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTCGAGCCGGCAACGG 2685

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| |||||||||||  |||||||||||||
Db       2686 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2745

Qy        601 ATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||||||||||||||| |||||||||||||||||
Db       2746 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 2780


RESULT 18
US-09-902-535-1
; Sequence 1, Application US/09902535
; Patent No. US20020102530A1
; GENERAL INFORMATION:
;  APPLICANT: Keith, Jr., James C.
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  Mi, Sha
;  TITLE OF INVENTION: Methods and compositions for diagnosing
;  TITLE OF INVENTION:  and treating preeclampsia and gestational trophoblast
;  TITLE OF INVENTION:  disorders
;  FILE REFERENCE: GIN-6006B4
;  CURRENT APPLICATION NUMBER: US/09/902,535
;  CURRENT FILING DATE:  2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,657
;  PRIOR FILING DATE: 2000-07-06
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2930
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (930)...(2546)
US-09-902-535-1

  Query Match              81.5%;  Score 517.4;  DB 9;  Length 2930;
  Best Local Similarity   91.4%;  Pred. No. 1.7e-150;
  Matches  564;  Conservative    0;  Mismatches   41;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2325 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2384

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2385 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2432

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2433 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2492

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2493 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2552
```

```
Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2553 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2612

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       2613 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2672

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2673 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2732

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       2733 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2792

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2793 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2852

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
              ||||||||||||||||||||||||||||||||| |||||||||||  |||||||||||||
Db       2853 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 2912

Qy        601 ATTAAATCTTGCAACTG 617
              |||||||||||||||||
Db       2913 ATTAAATCTTGCAACTG 2929


RESULT 19
US-09-873-367C-81
; Sequence 81, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  APPLICANT:  Soppet, Daniel
;  APPLICANT:  Endress, Gregory
;  APPLICANT:  Augustus, Meena
;  APPLICANT:  Ebner, Reinhard
;  APPLICANT:  Carter, Kenneth
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
;  TITLE OF INVENTION:  Signature Gene Sets
;  FILE REFERENCE: 689290-64
;  CURRENT APPLICATION NUMBER: US/09/873,367C
;  CURRENT FILING DATE:  2003-04-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,891
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,842
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/244,867
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: U.S. 60/245,084
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 1067
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 81
;   LENGTH: 56093
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-873-367C-81

  Query Match              81.3%;  Score 516.2;  DB 10;  Length 56093;
  Best Local Similarity   91.1%;  Pred. No. 1.6e-149;
  Matches  564;  Conservative    0;  Mismatches   43;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db      37274 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 37333

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db      37334 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 37381
```

```
Qy         121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
               ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       37382 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 37441

Qy         181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
               ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       37442 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 37501

Qy         241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
               | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       37502 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 37561

Qy         301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
               |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       37562 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 37621

Qy         361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       37622 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 37681

Qy         421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
               |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db       37682 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 37741

Qy         481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
               |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       37742 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 37801

Qy         541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
               ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       37802 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 37861

Qy         601 ATTAAATCTTGCAACTGAA 619
               ||||||||||||||||| |
Db       37862 ATTAAATCTTGCAACTGCA 37880


RESULT 20
US-10-632-793-30
; Sequence 30, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
;   LENGTH: 7582
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (307)..(307)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (355)..(355)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1309)..(1309)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1331)..(1331)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2213)..(2213)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2398)..(2398)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3787)..(3787)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4115)..(4115)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4261)..(4261)
;   OTHER INFORMATION: n = a or g or c or t/u
US-10-632-793-30

  Query Match             80.5%;  Score 511.4;  DB 16;  Length 7582;
  Best Local Similarity   89.5%;  Pred. No. 2e-148;
  Matches  552;  Conservative   11;  Mismatches   42;  Indels   12;  Gaps    1;

Qy           1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
               ||||||||||||:||||||||||||| |||||||||||||||||| || ||||:||||||
Db        6976 CCCTGTATCTTTRACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTRAAACTA 7035

Qy          61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                           |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        7036 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 7083

Qy         121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
               |||||||||||||:||||||| | || ||||||||||||||  ||||||||||||| |||
Db        7084 CTGGACCGGCCTGYTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 7143

Qy         181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
               ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        7144 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 7203

Qy         241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
               : || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db        7204 SGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 7263

Qy         301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
               |||||||||||||||||||||||  |: ||||||||:|||||||||:||||||||||||
Db        7264 AGGACTAGCTGGATTTCCTAGGCTGAYTAAGAATCCYTAAGCCTAGSTGGGAAGGTGACC 7323

Qy         361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                ||||||||| |||||| ||||||||||||||||:||||||| |||||||||||||||||
Db        7324 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGYTCACACCTGACCAATCAGAGAGCTC 7383

Qy         421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
               |||||||||||||||||||||| | |||||||||||||||||||||||||:||||||:||
Db        7384 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATYTATTGCMTG 7443

Qy         481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
               |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db        7444 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTYTTCGAGCCGGCAACGG 7503

Qy         541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600
               ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||||
Db       7504 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCT 7563

Qy        601 ATTAAATCTTGCAACTG 617
              |||||||||||||:|||
Db       7564 ATTAAATCTTGCARCTG 7580

RESULT 21
US-10-087-192-730/c
; Sequence 730, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 730
;   LENGTH: 161334
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(161334)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-730

  Query Match             80.4%;  Score 510.6;  DB 13;  Length 161334;
  Best Local Similarity   90.0%;  Pred. No. 1.5e-147;
  Matches  575;  Conservative    0;  Mismatches   49;  Indels   15;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||  |||||| |||
Db     100611 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGTTGTAAAGCTA 100552

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||| | ||||||| ||||||||||||||||| |||||||||||
Db     100551 CAAATCGTTCTTCAAATGGAACCCCAGATGAAGTCCATGACTAAGATCTACCGTGGACCC 100492

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||||||| |||||||||||||| ||  |||||| || ||| ||||||||||||||
Db     100491 CTGGACCGGCCTACTAGCCCATGCTCCAATTGTAATGATATCGAACGCACCCCTCCCGAG 100432

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||||||||||||||||||||| || |||||||||||||  |||
Db     100431 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCCGCAGGAAGCAGTTAGACTGGT 100372

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     100371 CGTCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGTGGGGGACTGAGAGAC 100312

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||| ||||||||||||||||||| || ||||||||| |||||||||||||||||||||
Db     100311 AGGATTAGCTGGATTTCCTAGGCCGACTAAGAATCCCAAAGCCTAGCTGGGAAGGTGACC 100252

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
               ||||||||| ||||||  ||||||||||||||||||||||||||||||||
Db     100251 ACATCCACCTTTAAACACTGGGCTTGCAACTTAGCTCACACCCGACCAATCAGGTAGTAA 100192

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              |||||||||||||||||||||||||| |||||| |||||||||| |||||||||||||||
Db     100191 AGAGAGCTCACTAAAATGCTAATTAGACAAAAACAGGAGGTAAAAAAATAGCCAATCATC 100132

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              ||| ||||||||||||||||||| |||||| ||||||||||||||||||||||||| |||
Db     100131 TATCGCCTGAGAGCACAGCGGGAAGGACAATGATCGGGATATAAACCCAGGCATTCAAGC 100072

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              |||||||||| |||  |||||||||||||||||||||||||| ||||| | |
Db     100071 CGGCAACGGCTACCTTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTCTCT------GT 100018

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAA 630
               ||||||||||||||| ||||||||| |||||||||| |
Db     100017 CTTCACTCTATTAAATATTGCAACTGCAAAAAAAAAAATA 99979

RESULT 22
US-10-087-192-1666
; Sequence 1666, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1666
;   LENGTH: 285020
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(285020)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1666

  Query Match             78.9%;  Score 500.8;  DB 13;  Length 285020;
  Best Local Similarity   89.3%;  Pred. No. 2.3e-144;
  Matches  577;  Conservative    0;  Mismatches   57;  Indels   12;  Gaps    3;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||||||||||||||||||||||||||||||||||||||||| || |||||||||||
Db     279486 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 279545

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| |||||||||||||||| ||| ||||||||
Db     279546 CAAATAGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATTCACTGTGGACCC 279605

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||||||||||||||| |||||||  ||||||||||||| |||
Db     279606 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTGATGACATCAAAGGCACCCCTCCTGAG 279665

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||| |||||||||||| |||||||||||||||||||||||| ||||||||||||||| |||
Db     279666 GAACTCTCAACTGCACGACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGT 279725

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |||  |||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db     279726 CATTGGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGTGGGGACTGAGAGAC 279785

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAG--AATCCCTAAGCCTAGCTGGGAAGGTGA 358
              |||||||||||||||||||||||  || |||  |||||||||||||||||||||||||||
Db     279786 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAAAATCCCTAAGCCTAGCTGGGAAGGTGA 279845

Qy        359 CTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC------- 411
              | |||||||||| |||||| |||||||||||||||||||||||||||||||||
Db     279846 CGGCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGGGAGT 279905

Qy        412 --AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCA 469
                ||| |||||||||||||||||||||||| |||| ||||||||||||||||||||||||
Db     279906 AAAGACAGCTCACTAAAATGCTAATTAGGCTAAAACAGGAGGTAAAGAAATAGCCAATCA 279965

Qy        470 TCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGA 529
              ||||| |||||||||||| | ||||||||||| |||| |||||||||||||||||||| |
Db     279966 TCTATCGCCTGAGAGCAC-GAGGGAGGGACAATGATCAGGATATAAACCCAGGCATTCAA 280024

Qy        530 GCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTC 589
              |||||||  ||| |||| ||||||||||||||  |||||||||| |||||||||||||||
Db     280025 GCCGGCAGTGGCTACCCGCTTTGGGTCCCCTCTGTTTGTATGGGAGCTCTGTTTTCACTC 280084

Qy        590 TATTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              ||||  |   |   |   |    ||   ||||||||||||||||||
Db     280085 TATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAA 280130


RESULT 23
US-09-864-761-4444
; Sequence 4444, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4444
;   LENGTH: 1894
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC002346.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =   6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444

  Query Match             78.7%;  Score 499.6;  DB 9;  Length 1894;
  Best Local Similarity   88.7%;  Pred. No. 5.2e-145;
  Matches  571;  Conservative    0;  Mismatches   54;  Indels   19;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||| |||||||||| || | |||||||||
Db       1217 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAATCGAAGCAGTAAAACTA 1276

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||| ||| ||| |||  ||||||
Db       1277 CAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAGTAAAATCTACCACGGACCC 1336

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              |||||||||||||||||||||||||| ||||||||||||||  |||||||||||||||||
Db       1337 CTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAAAGGCACCCCTCCCGAG 1396

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||||||||||||| |||||| ||||||||||||| |||||||||||||| |||
Db       1397 GAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGT 1456

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |  ||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       1457 TGTTGGCCAACCTCCCCAACAGCAGTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 1516

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||| || || |||||||||| ||||| ||||||||||||| |||||||||||||||||
Db       1517 AGGAATAACTAGATTTCCTAGACCAACTAAGAATCCCTAAGACTAGCTGGGAAGGTGACC 1576

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
              || ||||||| ||||||  |||||||||||||||||||| ||| |||||||
Db       1577 GCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCAACCAATCAGATACTAA 1636

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              ||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||||||
Db       1637 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGATAAAGAAATAGCCAATCATC 1696

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              | |||||||| |||||||| |||||||||| |||||||||||||||||||||||||||||
Db       1697 TGTTGCCTGACAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCATTCGAGC 1756

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              | || || || |||| |||||||||||||||||||||||||| |||||||
Db       1757 CAGCTACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT---------- 1806

Qy        592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
               ||||||||||||||||||||||||| ||||| ||||| | |||
Db       1807 CTTCACTCTATTAAATCTTGCAACTGCAAAAATAAAAATAGAAA 1850


RESULT 24
US-10-632-793-28
; Sequence 28, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
;   LENGTH: 3372
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-632-793-28

  Query Match             78.0%;  Score 495.6;  DB 16;  Length 3372;
  Best Local Similarity   92.2%;  Pred. No. 1.2e-143;
  Matches  522;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy         70 CTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGG 129
              || |||||||||| | ||||| |||||| |||||||||| ||||  ||||||||||||||
Db       2802 CTACAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGGACCGG 2861

Qy        130 CCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCA 189
              |||| ||||||| | || ||||||||||||||  ||||||||||||| ||||||||||||
Db       2862 CCTGTTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAATCTCA 2921

Qy        190 ACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCA 249
               |||||||||| |||||| ||||||||||||| ||||||||||||||||||| || ||||
Db       2922 GCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTCGGCCA 2981

Qy        250 ACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGC.309
              |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||
Db       2982 ACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGACAGGACTAGC 3041

Qy        310 TGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACC 369
              ||||||||||||||  |  |||||||||||||||||||||||||||||||  ||||||||
Db       3042 TGGATTTCCTAGGCTGATTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACATCCACC 3101

Qy        370 TCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAAATG 429
              | |||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       3102 TTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATG 3161

Qy        430 CTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAG 489
              ||||||||||||| | ||||||||||||||||||||||||| ||||||||||||||||||
Db       3162 CTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATTTATTGCCTGAGAGCACAG 3221

Qy        490 CGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCT 549
              | |||||||||| |||||||||||||||||| |  |||||||||||||||||||||||||
Db       3222 CAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTCGAGCCGGCAACGGCAACCCCCT 3281

Qy        550 TTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCT 609
              |||||||||||||||||||||||| ||||||||||||  |||||||||||||||||||||
Db       3282 TTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCACTCTATTAAATCT 3341

Qy        610 TGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||| |||||||||||||||||
Db       3342 TGCAACTGCAAAAAAAAAAAAAAAAA 3367

RESULT 25
US-10-276-774-678
; Sequence 678, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  APPLICANT:  Tang, Y, Tom et al
;  TITLE OF INVENTION: No. US20040053245Alel Nucleic Acids and Polypeptides
;  FILE REFERENCE: 21272-030
;  CURRENT APPLICATION NUMBER: US/10/276,774
;  CURRENT FILING DATE:  2002-11-18
;  PRIOR APPLICATION NUMBER: 09/560,875
;  PRIOR FILING DATE: 2000-04-27
```

```
;  PRIOR APPLICATION NUMBER: 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  NUMBER OF SEQ ID NOS: 2700
;  SOFTWARE: Custom
; SEQ ID NO 678
;   LENGTH: 2052
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-276-774-678

  Query Match             77.7%;  Score 493.2;  DB 16;  Length 2052;
  Best Local Similarity   91.2%;  Pred. No. 5.3e-143;
  Matches  549;  Conservative    0;  Mismatches   43;  Indels   10;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |
Db        993 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACAA 1052

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              || || |||||||||||||||| ||||||| ||||||||||||||||| |||| ||| ||
Db       1053 CACATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCGCGGATCC 1112

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||| ||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db       1113 CTGGACCAGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACTCCTCCCGAG 1172

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              |||||||||| |||||||||||||||||||||||||| ||| |||||||||| |||| ||
Db       1173 GAAATCTCAAATGCACAACCCCTACTATGCCCCAATTAAGCAGGAAGCAGTTGGAGCAGT 1232

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              |  | |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1233 CGACGGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAGAGAC 1292

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              ||||||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db       1293 AGGACTAGCTGGATTTCCTAGGCCGGATAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 1352

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 411
              | |||||||| ||||||  ||||||||||||||||||||||||||||||||
Db       1353 GTATCCACCTTTAAACACAGGGCTTGCAACTTAGCTCACACCCGACCAATCAGGTAGTAA 1412

Qy        412 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC 471
              ||||||||||||||||||||||||||||||||| |||||||||||||||||| ||| |||
Db       1413 AGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCTAAT-ATC 1471

Qy        472 TATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGC 531
              |||  |||||||| |||| ||||||||||| ||| |||||| ||||||||||||||||||
Db       1472 TATCACCTGAGAGTACAGGGGGAGGGACAATGATTGGGATAGAAACCCAGGCATTCGAGC 1531

Qy        532 CGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTA 591
              ||||||||||||||||||||||||| ||| || || |||||| |||||||||||||||||
Db       1532 CGGCAACGGCAACCCCCTTTGGGTCTCCTTCCATTTTATGGGAGCTCTGTTTTCACTCTA 1591

Qy        592 TT 593
              ||
Db       1592 TT 1593.

RESULT 26
US-09-997-722-148
; Sequence 148, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David
;  APPLICANT:  Engelhard, Eric
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: A-71171/RMS/DCF
;  CURRENT APPLICATION NUMBER: US/09/997,722
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 301
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 148
;   LENGTH: 22436
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-997-722-148

  Query Match            76.0%;  Score 482.6;  DB 11;  Length 22436;
  Best Local Similarity  87.6%;  Pred. No. 3.4e-139;
  Matches  565;  Conservative    0;  Mismatches   69;  Indels   11;  Gaps    3;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              ||||||| ||||||||||||||||||||||||||||||||||||| || |||||||||||
Db       5655 CCCTGTACCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 5714

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
              ||||| |||||||||||||||| ||||||| ||||||||||||||||| || |   ||||
Db       5715 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACTGCAGACCC 5774

Qy        121 CTGGACCGGCCTGCTAGCC-CATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGA 179
              |||||| |||||||||||| |||||||| ||||||||||||||||||||| | |||||||
Db       5775 CTGGACTGGCCTGCTAGCCTCATGCTCCAATGTTAATGACATTGAAGGCAACTCTCCCGA 5834

Qy        180 GGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGG 239
              ||||||||||||||||||||||||||||  ||||| |||||| |||||||||||||||
Db       5835 GGAAATCTCAACTGCACAACCCCTACTACACCCCAGTTCAGCAGGAAGCAGTTAGAGCAA 5894

Qy        240 TCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGA 299
              ||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5895 TCATTGGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAG- 5953

Qy        300 CAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 359
              ||||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       5954 CAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 6013

Qy        360 TGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC-------- 411
              |||| |||||| |||||| |||||||||||||||| ||||||| ||||||||
Db       6014 TGCACCCACCTTTAAACACGGGGCTTGCAACTTAGTTCACACCTGACCAATCAAGTAGTA 6073

Qy        412 -AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCAT 470
               ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db       6074 AAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCAT 6133

Qy        471 CTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAG 530
              ||||  ||||||||||||||| | ||||||| |||| ||||||||||||||||||| |||
Db       6134 CTATCACCTGAGAGCACAGCGTGGGGGACAATGATCAGGATATAAACCCAGGCATTTGAG 6193

Qy        531 CCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCT 590
              || |||| ||| ||||| ||||||||||||||||||||||||| ||||||||| ||||||
Db       6194 CCAGCAATGGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTACACTCT 6253

Qy        591 ATTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||  ||  |   |   |     |   | | | | |  ||||| |
Db       6254 ATTAAACCTTGCAACTGCAAAATATATATATATATATGAAAAATA 6298

RESULT 27
US-10-220-120-15
; Sequence 15, Application US/10220120
; Publication No. US20040048253A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  PANZER, Scott R.
;  APPLICANT:  SPIRO, Peter A.
;  APPLICANT:  BANVILLE, Steven C.
;  APPLICANT:  SHAH, Purvi
;  APPLICANT:  CHALUP, Michael S.
;  APPLICANT:  CHANG, Simon C.
;  APPLICANT:  CHEN, Alice
;  APPLICANT:  D'SA, Steven A.
;  APPLICANT:  AMSHEY, Stefan
```

```
;  APPLICANT:  DAHL, Christopher R.
;  APPLICANT:  DAM, Tam C.
;  APPLICANT:  DANIELS, Susan E.
;  APPLICANT:  DUFOUR, Gerard E.
;  APPLICANT:  FLORES, Vincent
;  APPLICANT:  FONG, Willy T.
;  APPLICANT:  GREENAWALT, Lila B.
;  APPLICANT:  HILLMAN, Jennifer L.
;  APPLICANT:  JONES, Anissa L.
;  APPLICANT:  LIU, Tommy F.
;  APPLICANT:  ROSEBERRY, Ann M.
;  APPLICANT:  ROSEN, Bruce H.
;  APPLICANT:  RUSSO, Frank D.
;  APPLICANT:  STOCKDREHER, Theresa K.
;  APPLICANT:  DAFFO, Abel
;  APPLICANT:  WRIGHT, Rachel J.
;  APPLICANT:  YAP, Pierre E.
;  APPLICANT:  YU, Jimmy Y.
;  APPLICANT:  BRADLEY, Diana L.
;  APPLICANT:  BRATCHER, Shawn R.
;  APPLICANT:  CHEN, Wensheng
;  APPLICANT:  COHEN, Howard J.
;  APPLICANT:  HODGSON, David M.
;  APPLICANT:  LINCOLN, Stephen E.
;  APPLICANT:  JACKSON, Stuart
;  TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PT-1113 PCT
;  CURRENT APPLICATION NUMBER: US/10/220,120
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774;
;    60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;
;    60/184,769; 60/184,768; 60/184,837; 60/184,697; 60/184,841;
;    60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;
;    60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;
;    60/204,525; 60/205,285; 60/205,232; 60/205,323; 60/205,287;
;    60/205,324; 60/205,286
;  PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-05-17;
;    2000-05-17; 2000-05-12; 2000-05-16; 2000-05-16; 2000-05-15;
;    2000-05-16; 2000-05-17; 2000-05-16; 2000-05-17; 2000-05-17;
;    2000-05-17; 2000-05-17
;  NUMBER OF SEQ ID NOS: 422
;  SOFTWARE: PERL Program
; SEQ ID NO 15
;   LENGTH: 849
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20040048253A1 LG:429446.1:2000FEB18
US-10-220-120-15

  Query Match            75.8%;  Score 481.4;  DB 16;  Length 849;
  Best Local Similarity  89.1%;  Pred. No. 1.7e-139;
  Matches  534;  Conservative    0;  Mismatches   56;  Indels    9;  Gaps    1;

Qy         27 GTTTGTCTCTTCCAGAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATGGAGCACCA 86
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        121 GTTTGTCTCTTCCAGAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATGGAGCCCCA 180

Qy         87 GATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAGCCCATGCTC 146
              |||| |||| |||||||||||| |||   ||||| |||||||||||||||||||||||||
Db        181 GATGCAGTCTATGACTAAGATCTACCACAGACCCTTGGACCGGCCTGCTAGCCCATGCTC 240

Qy        147 CGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACT 206
              ||||||| ||||||| |||||| |||||||| |||||||||||||||||  |||||||||
Db        241 CGATGTTGATGACATCGAAGGCGCCCCTCCCAAGGAAATCTCAACTGCATGACCCCTACT 300

Qy        207 ATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCACT 266
              | |||||||||| || ||||||||||||||| |||||   ||||||||||||||||||||
Db        301 ACGCCCCAATTCCGCAGGAAGCAGTTAGAGCAGTCATTGGCCAACCTCCCCAACAGCACT 360
```

```
Qy        267 TGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAA 326
              || |||||||||||||||||| | |||||||||||||||||||||||||||| |||||||
Db        361 TGAGTTTTCCTGTTGAGAGGGAGAACTGAGAGACAGGACTAGCTGGATTTCCCAGGCCAA 420

Qy        327 CGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTG 386
              | ||||||||||||||||| ||||||||||||| | |||||||| ||||||| |||||||
Db        421 CTAAGAATCCCTAAGCCTATCTGGGAAGGTGACCGTATCCACCTTTAAACATAGGGCTTG 480

Qy        387 CAACTTAGCTCACACCCGACCAATCA---------GAGAGCTCACTAAAATGCTAATTAG 437
              |||||||||||||||||||| |||||         |||||||||||||||||||||||||
Db        481 CAACTTAGCTCACACCCGACTAATCAGATAGTAAGGAGAGCTCACTAAAATGCTAATTAG 540

Qy        438 GCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGG 497
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        541 GCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCAGGAGGG 600

Qy        498 ACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCC 557
              |||| |||| |||||||||||||||||||| |||| |||| ||| |||| ||||||||||
Db        601 ACAATGATCAGGATATAAACCCAGGCATTCAAGCCAGCAATGGCTACCCTCTTTGGGTCC 660

Qy        558 CCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCTTGCAACT 616
              ||||| |||||||||| ||||||||||||  |||||| |   | ||  | ||    |||
Db        661 CCTCCGTTTGTATGGGAGCTCTGTTTTCATGCTATTTTAAAGTCTTGCAACTGCACACT 719

RESULT 28
US-10-104-047-1148
; Sequence 1148, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1148
;   LENGTH: 2349
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-104-047-1148

  Query Match             73.4%;  Score 466;  DB 15;  Length 2349;
  Best Local Similarity   86.4%;  Pred. No. 1.8e-134;
  Matches  541;  Conservative    0;  Mismatches   75;  Indels   10;  Gaps    2;

Qy         19 CTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATG 78
              ||| || ||||||||||||||||||||||| || ||||||||||||| ||||| ||||||
Db       1377 CTTCTTCAGTTTGTCTCTTCCAGAATCAAAGCTATAAAACTACAAATGGTTCTCCAAATG 1436

Qy         79 GAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAGC 138
              |||| ||||||| ||||||||||||||||| || | ||||| ||||||||||||||||||
Db       1437 GAGCCCCAGATGCAGTCCATGACTAAGATCTACTGCGGACCTCTGGACCGGCCTGCTAGC 1496

Qy        139 CCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTGCACAA 198
              ||| ||||| |  ||||||||||  ||||||||||||||||||||||||||||||||  |
Db       1497 CCA-GCTCCCACATTAATGACATCAAAGGCACCCCTCCCGAGGAAATCTCAACTGCATGA 1555

Qy        199 CCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCA 258
              | ||||||| ||||||||||||| |||||||||||||| |||| |  |||||||||||||
Db       1556 CACCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGTCGTTGGCCAACCTCCCCA 1615

Qy        259 ACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCC 318
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1616 ACAGCACTTGGGTTTTCCTGTTGAGATGGGGGACTGAGAGACAGGACTAGCTGGATTTCC 1675

Qy        319 TAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACAT 378
              ||||||||| |||||||||||||||||||||||||||||||| || |||||| ||| ||
Db       1676 TAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTACACCCACCTTTAACCAC 1735
```

```
Qy        379 GGGGCTTGCAACTTAGCTCACACCCGACCAATC---------AGAGAGCTCACTAAAATG 429
                |||||||||||||||||||||||||||||||         ||||||||  ||||||||
Db       1736 TAGGCTTGCAACTTAGCTCACACCCGACCAATCAGGTAGTAAAGAGAGCTTGCTAAAATG 1795

Qy        430 CTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAG 489
              ||||||||||||||| |||||||||||||||||||| |||||||| |||||| ||||||
Db       1796 CTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAGTCATCTATCGCCTGACAGCACAA 1855

Qy        490 CGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCT 549
               ||| ||||||| |||| |||||||||| ||||||||| |||| |||| ||| |||| ||
Db       1856 GGGGCGGGACAATGATCAGGATATAAACTCAGGCATTCAAGCCAGCAATGGCTACCCACT 1915

Qy        550 TTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCT 609
              |||||||||||||| || |||||| |||||||||||||||||||  |   |   |   |
Db       1916 TTGGGTCCCCTCCCATTTTATGGGAGCTCTGTTTTCACTCTATTAAATCTTTCAACTACC 1975

Qy        610 TGCAACTGAAAAAAAAAAAAAAAAAA 635
                 ||   ||||||| |||| |||||
Db       1976 AAAAAAAAAAAAAAAGAAAAGAAAAA 2001

RESULT 29
US-10-204-887-10
; Sequence 10, Application US/10204887
; Publication No. US20030124569A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  PANZER, Scott R.
;  APPLICANT:  SPIRO, Peter A.
;  APPLICANT:  BANVILLE, Steven C.
;  APPLICANT:  SHAH, Purvi
;  APPLICANT:  CHALUP, Michael S.
;  APPLICANT:  CHANG, Simon C.
;  APPLICANT:  CHEN, Alice
;  APPLICANT:  D'SA, Steven A.
;  APPLICANT:  AMSHEY, Stefan
;  APPLICANT:  DAHL, Christopher R.
;  APPLICANT:  DAM, Tam C.
;  APPLICANT:  DANIELS, Susan E.
;  APPLICANT:  DUFOUR, Gerard E.
;  APPLICANT:  FLORES, Vincent
;  APPLICANT:  FONG, Willy T.
;  APPLICANT:  GREENAWALT, Lila B.
;  APPLICANT:  HILLMAN, Jennifer L.
;  APPLICANT:  JONES, Anissa L.
;  APPLICANT:  LIU, Tommy F.
;  APPLICANT:  ROSEBERRY, Ann M.
;  APPLICANT:  ROSEN, Bruce H.
;  APPLICANT:  RUSSO, Frank D.
;  APPLICANT:  STOCKDREHER, Theresa K.
;  APPLICANT:  DAFFO, Abel
;  APPLICANT:  WRIGHT, Rachel J.
;  APPLICANT:  YAP, Pierre E.
;  APPLICANT:  YU, Jimmy Y.
;  APPLICANT:  BRADLEY, Diana L.
;  APPLICANT:  BRATCHER, Shawn R.
;  APPLICANT:  CHEN, Wensheng
;  APPLICANT:  COHEN, Howard J.
;  APPLICANT:  HODGSON, David M.
;  APPLICANT:  LINCOLN, Stephen E.
;  TITLE OF INVENTION: SECRETORY MOLECULES
;  FILE REFERENCE: PT-1134 PCT
;  CURRENT APPLICATION NUMBER: US/10/204,887
;  CURRENT FILING DATE:  2002-08-21
;  PRIOR APPLICATION NUMBER: 60/185,215; 60/185,216; 60/205,232; 60/205,323; 60/205,287;
;    60/205,324; 60/205,286
;  PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;
;    2000-05-17; 2000-05-17
;  NUMBER OF SEQ ID NOS: 159
;  SOFTWARE: PERL Program
; SEQ ID NO 10
;   LENGTH: 1564
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:813218.1:2000FEB01
;  FEATURE:
;  NAME/KEY: unsure
;  LOCATION: 307, 317-318, 335, 918, 921, 943, 1011
;  OTHER INFORMATION: a, t, c, g, or other
US-10-204-887-10

  Query Match             70.7%;  Score 449;  DB 15;  Length 1564;
  Best Local Similarity   90.9%;  Pred. No. 3e-129;
  Matches  490;  Conservative    0;  Mismatches   45;  Indels    4;  Gaps    1;

Qy         41 GAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGA 100
              ||||  ||||||||||||||||||| |||| ||||||||||| ||||||| |||||||||
Db       1022 GAATTGAAACTGTAAAACTACAAATGGTTCATCAAATGGAGCCCCAGATGCAGTCCATGA 1081

Qy        101 CTAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACA 160
              |||||||||||||| |||||| ||||||| || | ||||||||||| | |||||||||||
Db       1082 CTAAGATCCACCGTAGACCCCCGGACCGGTCTCCCAGCCCATGCTCTGGTGTTAATGACA 1141

Qy        161 TTGAAGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAG 220
              | ||||||||||||||| ||||||||||| ||||||||||||| |||||||||| |||||
Db       1142 TCGAAGGCACCCCTCCCAAGGAAATCTCAGCTGCACAACCCCTCCTATGCCCCAGTTCAG 1201

Qy        221 CGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTT 280
              | ||||||||||||||| |||||| |||||||||||||| ||||||||||||||||||||
Db       1202 CAGGAAGCAGTTAGAGCAGTCATCGGCCAACCTCCCCAATAGCACTTGGGTTTTCCTGTT 1261

Qy        281 GAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAA 340
              ||||| ||||||||||    |||||||||||||||||||||||| || ||||||||||||
Db       1262 GAGAGTGGGGACTGAG----AGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAA 1317

Qy        341 GCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACA 400
              |||||||||||||||| ||| |||||| || |||||||||||||||||||||||||||||
Db       1318 GCCTAGCTGGGAAGGTAACTACATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACA 1377

Qy        401 CCCGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAAT 460
              ||| |||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db       1378 CCCAACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAAT 1437

Qy        461 AGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCA 520
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |||
Db       1438 AGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCAGGATATAAATCCA 1497

Qy        521 GGCATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCT 579
              ||||||| |||| |||| |||||||||||||||||| ||||| | | |  | || || |
Db       1498 GGCATTCCAGCCAGCAATGGCAACCCCCTTTGGGTCGCCTCCTTGTATGGAGCTCTAT 1556


RESULT 30
US-10-220-120-17
; Sequence 17, Application US/10220120
; Publication No. US20040048253A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  PANZER, Scott R.
;  APPLICANT:  SPIRO, Peter A.
;  APPLICANT:  BANVILLE, Steven C.
;  APPLICANT:  SHAH, Purvi
;  APPLICANT:  CHALUP, Michael S.
;  APPLICANT:  CHANG, Simon C.
;  APPLICANT:  CHEN, Alice
;  APPLICANT:  D'SA, Steven A.
;  APPLICANT:  AMSHEY, Stefan
;  APPLICANT:  DAHL, Christopher R.
;  APPLICANT:  DAM, Tam C.
;  APPLICANT:  DANIELS, Susan E.
;  APPLICANT:  DUFOUR, Gerard E.
;  APPLICANT:  FLORES, Vincent
;  APPLICANT:  FONG, Willy T.
;  APPLICANT:  GREENAWALT, Lila B.
;  APPLICANT:  HILLMAN, Jennifer L.
;  APPLICANT:  JONES, Anissa L.
;  APPLICANT:  LIU, Tommy F.
;  APPLICANT:  ROSEBERRY, Ann M.
;  APPLICANT:  ROSEN, Bruce H.
;  APPLICANT:  RUSSO, Frank D.
;  APPLICANT:  STOCKDREHER, Theresa K.
;  APPLICANT:  DAFFO, Abel
;  APPLICANT:  WRIGHT, Rachel J.
;  APPLICANT:  YAP, Pierre E.
;  APPLICANT:  YU, Jimmy Y.
;  APPLICANT:  BRADLEY, Diana L.
;  APPLICANT:  BRATCHER, Shawn R.
;  APPLICANT:  CHEN, Wensheng
;  APPLICANT:  COHEN, Howard J.
;  APPLICANT:  HODGSON, David M.
;  APPLICANT:  LINCOLN, Stephen E.
;  APPLICANT:  JACKSON, Stuart
;  TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PT-1113 PCT
;  CURRENT APPLICATION NUMBER: US/10/220,120
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774;
;    60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;
;    60/184,769; 60/184,768; 60/184,837; 60/184,697; 60/184,841;
;    60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;
;    60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;
;    60/204,525; 60/205,285; 60/205,232; 60/205,323; 60/205,287;
;    60/205,324; 60/205,286
;  PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-05-17;
;    2000-05-17; 2000-05-12; 2000-05-16; 2000-05-16; 2000-05-15;
;    2000-05-16; 2000-05-17; 2000-05-16; 2000-05-17; 2000-05-17;
;    2000-05-17; 2000-05-17
;  NUMBER OF SEQ ID NOS: 422
;  SOFTWARE: PERL Program
; SEQ ID NO 17
;   LENGTH: 1393
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20040048253A1 LI:351965.1:2000FEB01
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 120, 765
;   OTHER INFORMATION: a, t, c, g, or other
US-10-220-120-17

  Query Match             70.3%;  Score 446.2;  DB 16;  Length 1393;
  Best Local Similarity   90.3%;  Pred. No. 2.2e-128;
  Matches  502;  Conservative    0;  Mismatches   44;  Indels   10;  Gaps    2;

Qy         47 AAACTGTAAAACTACAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGA 106
              || |||||||||||| |||||||||||||| ||||| ||||||| |||||||||||||||
Db          2 AAGCTGTAAAACTACTAATTGTTCTTCAAAAGGAGCCCCAGATGCAGTCCATGACTAAGA 61

Qy        107 TCCACCGTGGACCCCTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAG 166
              || |||   ||||||||||||||||||||||||||||| ||||||||||||| || ||
Db         62 TCTACCACAGACCCCTGGACCGGCCTGCTAGCCCATGCACCGATGTTAATGATATCGANA 121

Qy        167 GCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAA 226
              |||| || ||| ||||||| |||||||||||||||||||||  |||||||||||| ||||
Db        122 GCACTCC-CCCAAGGAAATTTCAACTGCACAACCCCTACTACACCCCAATTCAGCAGGAA 180

Qy        227 GCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGG 286
              ||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||||
Db        181 GCAGTTAGAGCGGTCGTCAGCCAACCTCCCCAGCAGCACTTGGGTTTTCCTGTTGAGAGC 240

Qy        287 GGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAG 346
```

```
              |||||||||||||||||||||||||||||||||||||  || ||||||||||||||||||
Db        241 GGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAG 300

Qy        347 CTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGAC 406
              ||||||||||||| || || |||| |||||| ||||||||||||||||||||||||| ||
Db        301 CTGGGAAGGTGACCGCGTCTACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAAC 360

Qy        407 CAATC---------AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGA 457
              |||||         ||||||||||||||||||||||||||||||||| |||||||| |||
Db        361 CAATCAGGTAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAGAGA 420

Qy        458 AATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAAC 517
              ||||||||||||||||| ||||||||||||||| |||||||||| |||| ||||||||||
Db        421 AATAGCCAATCATCTATCGCCTGAGAGCACAGCAGGAGGGACAATGATCCGGATATAAAC 480

Qy        518 CCAGGCATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCT 577
              ||| ||||||||||| ||||||||  ||||||||| |||||||||||||||||||| |||
Db        481 CCAAGCATTCGAGCCAGCAACGGCTGCCCCCTTTGTGTCCCCTCCCTTTGTATGGGAGCT 540

Qy        578 CTGTTTTCACTCTATT 593
              ||||||||||||||||
Db        541 CTGTTTTCACTCTATT 556

Search completed: January 22, 2005, 20:08:26
Job time : 355.872 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 12:01:19 ; Search time 1891.54 Seconds
                                           (without alignments)
                                           12232.983 Million cell updates/sec

Title:          US-09-319-156B-6
Perfect score:  635
Sequence:       1 ccctgtatctttaacctcct..........tgaaaaaaaaaaaaaaaaaa 635

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters:      65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_htc:*
                4:  gb_est3:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_gss1:*
                9:  gb_gss2:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                    %
Result            Query
  No.    Score   Match Length DB  ID                   Description
-----------------------------------------------------------------------
     1   498.8    78.6   1160  3  BC030968             BC030968 Homo sapi
     2   492.8    77.6   2748  3  CR605851             CR605851 full-leng
     3   488.8    77.0   2749  3  CR617248             CR617248 full-leng
     4     488    76.9   2500  3  CR622175             CR622175 full-leng
     5   485.6    76.5    771  4  BI087886             BI087886 602852690
     6     485    76.4   2748  3  CR625046             CR625046 full-leng
     7   470.6    74.1    719  7  CN272394             CN272394 170006000
c    8   468.6    73.8   1071  5  BX365066             BX365066 BX365066
c    9   466.6    73.5    758  5  BX357208             BX357208 BX357208
    10   464.6    73.2   2716  3  CR613169             CR613169 full-leng
c   11   454.2    71.5    998  5  BX337769             BX337769 BX337769
c   12   453.6    71.4   1058  5  BX378303             BX378303 BX378303
c   13   450.8    71.0    689  9  AG121669             AG121669 Pan trogl
c   14     437    68.8    494  1  AA781423             AA781423 aj26c03.s
c   15     435    68.5   1019  5  BX439636             BX439636 BX439636
c   16   432.4    68.1    522  2  AW971553             AW971553 EST383642
    17   430.2    67.7    870  7  CN645411             CN645411 ILLUMIGEN
    18     424    66.8   1500  3  BC026287             BC026287 Homo sapi
c   19     423    66.6    653  9  AG033781             AG033781 Pan trogl
c   20   415.8    65.5    490  1  AI598135             AI598135 tn14a10.x
    21   397.4    62.6    440  2  BE732673             BE732673 601571305
c   22     396    62.4    543  4  BI963185             BI963185 ie57c10.x
c   23   395.2    62.2    966  5  BX380176             BX380176 BX380176
c   24   394.8    62.2    651  9  AG058970             AG058970 Pan trogl
c   25   393.2    61.9    701  9  AG126669             AG126669 Pan trogl
c   26     393    61.9    609  9  AG066901             AG066901 Pan trogl
c   27   392.6    61.8    443  1  AA837267             AA837267 od26b10.s
c   28   387.2    61.0    712  8  AQ892947             AQ892947 HS_3131_B
c   29   386.8    60.9    436  1  AI128526             AI128526 qc61h10.x
c   30   385.2    60.7    446  1  AI393478             AI393478 tg45g04.x
c   31   384.6    60.6    641  9  AG036829             AG036829 Pan trogl
c   32   384.2    60.5    485  2  AW511366             AW511366 hd45h03.x
c   33   379.8    59.8    664  5  BX481837             BX481837 DKFZp686M
c   34   375.2    59.1    556  1  AU158595             AU158595 AU158595
c   35   369.6    58.2    865  6  CB231128             CB231128 AGENCOURT
c   36   365.8    57.6    470  1  AI074704             AI074704 ox83d05.s
c   37   364.8    57.4    679  9  AG076758             AG076758 Pan trogl
c   38   364.2    57.4    415  1  AI128496             AI128496 qc61e08.x
c   39   360.6    56.8    777  9  AG030228             AG030228 Pan trogl
c   40   359.8    56.7    425  1  AI570707             AI570707 tm79g09.x
c   41   359.8    56.7    431  1  AA552941             AA552941 nk61a10.s
    42   359.2    56.6    490  1  AA426511             AA426511 zw02e05.r
c   43   359.2    56.6    619  9  AG133542             AG133542 Pan trogl
c   44   359.2    56.6    674  9  AG091649             AG091649 Pan trogl
c   45   355.4    56.0    446  6  CB069106             CB069106 is13a01.x

                                  ALIGNMENTS

RESULT 1
BC030968
LOCUS       BC030968                1160 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens cDNA clone IMAGE:4724433, with apparent retained
            intron.
ACCESSION   BC030968
VERSION     BC030968.1  GI:22658419
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1160)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1160)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:      http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 41 Row: m Column: 10
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1. .1160
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4724433"
                     /tissue_type="Placenta"
                     /clone_lib="NIH_MGC_79"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
ORIGIN

  Query Match             78.6%;  Score 498.8;  DB 3;  Length 1160;
  Best Local Similarity   92.6%;  Pred. No. 3.4e-133;
  Matches  524;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          70 CTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGG 129
               || |||||||||| | ||||| |||||| |||||||||| ||||  ||||||||||||||
Db         582 CTACAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGGACCGG 641

Qy         130 CCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCA 189
               |||||||||||| | || ||||||||||||||  ||||||||||||| ||||||||||||
Db         642 CCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAATCTCA 701

Qy         190 ACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCA 249
                |||||||||| |||||| ||||||||||||| ||||||||||||||||||| || ||||
Db         702 GCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTCGGCCA 761

Qy         250 ACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGC 309
               |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||
Db         762 ACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGACAGGACTAGC 821

Qy         310 TGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACC 369
               ||||||||||||||  || |||||||||||||||||||||||||||||||  ||||||||
Db         822 TGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACATCCACC 881

Qy         370 TCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAAATG 429
               | |||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db         882 TTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATG 941

Qy         430 CTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAG 489
               ||||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db         942 CTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAG 1001

Qy         490 CGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCT 549
               | |||||||||| |||||||||||||||||| |  |||||||||||||||||||||||||
Db        1002 CAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGGCAACCCCCT 1061

Qy         550 TTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCT 609
               |||||||||||||||||||||||| ||||||||||||  |||||||||||||||||||||
Db        1062 TTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCTATTAAATCT 1121

Qy         610 TGCAACTGAAAAAAAAAAAAAAAAAA 635
               |||||||  |||||||||||||||||
Db        1122 TGCAACTACAAAAAAAAAAAAAAAAA 1147


RESULT 2
CR605851
LOCUS       CR605851                 2748 bp    mRNA    linear   HTC 21-JUL-2004
```

```
DEFINITION  full-length cDNA clone CS0DE012YJ24 of Placenta of Homo sapiens
            (human).
ACCESSION   CR605851
VERSION     CR605851.1  GI:50486658
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2748)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2748)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .2748
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YJ24"
                     /tissue_type="Placenta"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             77.6%;  Score 492.8;  DB 3;  Length 2748;
  Best Local Similarity   90.9%;  Pred. No. 2.3e-131;
  Matches  540;  Conservative    0;  Mismatches   42;  Indels   12;  Gaps    1;

Qy           1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
               |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        2167 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2226

Qy          61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                           |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        2227 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2274

Qy         121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
               ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db        2275 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2334

Qy         181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
               ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        2335 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2394

Qy         241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
               | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db        2395 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2454

Qy         301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
               |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||| 
Db        2455 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2514

Qy         361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
                ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db        2515 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2574

Qy         421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
               |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db        2575 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2634
```

```
Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2635 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2694

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTT 594
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||||||
Db       2695 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTT 2748


RESULT 3
CR617248
LOCUS       CR617248                2749 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI022YJ18 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR617248
VERSION     CR617248.1  GI:50498055
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2749)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2749)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2749
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI022YJ18"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match            77.0%;  Score 488.8;  DB 3;  Length 2749;
  Best Local Similarity   90.8%;  Pred. No. 3.3e-130;
  Matches  536;  Conservative    0;  Mismatches   42;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2172 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2231

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2232 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2279

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2280 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2339

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2340 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2399

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2400 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGACTGAGAGAC 2459

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       2460 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGAAGGTGACC 2519

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2520 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2579

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db       2580 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2639

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2640 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2699

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCT 590
              ||||||||||||||||||||||||||||||||| ||||||||||||  ||
Db       2700 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCT 2749


RESULT 4
CR622175
LOCUS       CR622175                2500 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI051YM13 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR622175
VERSION     CR622175.1  GI:50502982
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2500)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2500)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2500
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match            76.9%;  Score 488;  DB 3;  Length 2500;
  Best Local Similarity   91.1%;  Pred. No. 5.5e-130;
  Matches  534;  Conservative    0;  Mismatches   40;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       1925 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 1984

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       1985 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2032
```

```
Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2033 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2092

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2093 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2152

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2153 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2212

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       2213 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2272

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2273 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2332

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db       2333 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2392

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2393 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2452

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCA 586
              ||||||||||||||||||||||||||||||||| ||||||||||||
Db       2453 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCA 2498
```

```
RESULT 5
BI087886
LOCUS       BI087886                 771 bp    mRNA    linear   EST 20-JUN-2001
DEFINITION  602852690F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4993894 5',
            mRNA sequence.
ACCESSION   BI087886
VERSION     BI087886.1  GI:14506216
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11015  row: d  column: 23
            High quality sequence stop: 762.
FEATURES             Location/Qualifiers
     source          1..771
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4993894"
                     /cell_line="MGC36"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_10"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5 kb. Library prepared by Life
                     Technologies."
```

```
ORIGIN

  Query Match             76.5%;  Score 485.6;  DB 4;  Length 771;
  Best Local Similarity   89.9%;  Pred. No. 2.1e-129;
  Matches  570;  Conservative    0;  Mismatches   49;  Indels   15;  Gaps    4;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        151 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 210

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        211 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 258

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db        259 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 318

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        319 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 378

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGT-TGAGAGGGGGGACTGAGAGA 299
              | || | |||||||||||| ||||||| ||||||||||| ||||| ||||||||||||||
Db        379 CGTC-GGCAACCTCCCCAAAAGCACTTAGGTTTTCCTGTGTGAGATGGGGGACTGAGAGA 437

Qy        300 CAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 359
              ||||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db        438 CAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 497

Qy        360 TGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCT 419
                ||||||||| |||||| |||||||||||||||||||||||| ||||||||||||||||
Db        498 CACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCT 557

Qy        420 CACTAAAATGCTAATTAGGCAAAAATAGGAGGTA-AAGAAATAGCCAATCATCTATTGCC 478
              ||||||||||||||||||||||| | |||||||| |||||||||||||||||||||||||
Db        558 CACTAAAATGCTAATTAGGCAAAGACAGGAGGTACAAGAAATAGCCAATCATCTATTGCC 617

Qy        479 TGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAAC 538
              |||||||||||| |||||||||| ||||||||||| || ||| |  ||||||||||||||
Db        618 TGAGAGCACAGCAGGAGGGACAATGATCGGGATATTAAACCAAGTCTTCGAGCCGGCAAC 677

Qy        539 GGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACT 598
              ||||||||||||||||||||||||||||||||||| ||||||| ||||  ||||||||||
Db        678 GGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTATTCATGCTATTTCACT 737

Qy        599 CTATTAAATCTTGCAACTGAAAAAAAAAAAAAAA 632
              ||||||||||||||||||| ||||||| |||| |
Db        738 CTATTAAATCTTGCAACTGCAAAAAAACAAAACA 771
```

```
RESULT 6
CR625046
LOCUS       CR625046              2748 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI044YK06 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR625046
VERSION     CR625046.1  GI:50505853
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2748)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2748)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2748
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI044YK06"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             76.4%;  Score 485;  DB 3;  Length 2748;
  Best Local Similarity   91.1%;  Pred. No. 4.1e-129;
  Matches  531;  Conservative    0;  Mismatches   40;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2178 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2237

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2238 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2285

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2286 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2345

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2346 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2405

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2406 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 2465

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db       2466 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 2525

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2526 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2585

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db       2586 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2645

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2646 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2705

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTT 583
              ||||||||||||||||||||||||||||||||| |||||||||
Db       2706 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTT 2748


RESULT 7
CN272394
LOCUS       CN272394                 719 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000600005410 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN272394
VERSION     CN272394.1  GI:47288808
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 719   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1..719
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cell, retinoic acid and
                     mitogen-treated hES cell line H7"
                     /clone_lib="GRN_PRENEU"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from hES cell line H7 (p29) maintained in feeder-free
                     conditions. Embryoid bodies were generated in the presence
                     of all-trans retinoic and mitogens."
ORIGIN

  Query Match             74.1%;  Score 470.6;  DB 7;  Length 719;
  Best Local Similarity   92.7%;  Pred. No. 4.7e-125;
  Matches  494;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         75 AATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCCTGC 134
              ||||||||| | ||||| |||||| |||||||||| ||||  ||||||||||||||||||
Db        187 AATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGGACCGGCCTGC 246

Qy        135 TAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTGC 194
              ||||||| | || ||||||||||||||  ||||||||||||| |||||||||||| ||||
Db        247 TAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAATCTCAGCTGC 306

Qy        195 ACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTC 254
              |||||| |||||| ||||||||||||| ||||||||||||||||||| || |||||||||
Db        307 ACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTCGGCCAACCTC 366

Qy        255 CCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGAT 314
              ||||||||||||| |||||||||||||||| |||||||||||||||||||||||||||||
Db        367 CCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGACAGGACTAGCTGGAT 426

Qy        315 TTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAA 374
              |||||||||  || |||||||||||||||||||||||||||||||  ||||||||| |||
Db        427 TTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACATCCACCTTTAA 486

Qy        375 ACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAAATGCTAAT 434
              ||| |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        487 ACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATGCTAAT 546

Qy        435 TAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGA 494
              |||||||| | ||||||||||||||||||||||||||||||||||||||||||||| |||
Db        547 TAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCAGGA 606

Qy        495 GGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCTTTGGG 554
              ||||||| |||||||||||||||||| |  ||||||||||||||||||||||||||||||
Db        607 GGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGGCAACCCCCTTTGGG 666

Qy        555 TCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAAT 607
              ||||||||||||||||||| ||||||||||||  |||||||||||||||||||
Db        667 TCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCTATTAAAT 719


RESULT 8
BX365066/c
```

```
LOCUS       BX365066                1071 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX365066 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI046YA18 3-PRIME, mRNA sequence.
ACCESSION   BX365066
VERSION     BX365066.2  GI:46304105
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1071)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30374869.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS1AI012ZC10NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..1071
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI046YA18"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             73.8%;  Score 468.6;  DB 5;  Length 1071;
  Best Local Similarity   89.9%;  Pred. No. 1.9e-124;
  Matches  525;  Conservative    3;  Mismatches   43;  Indels   13;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        571 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 512

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        511 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 464

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db        463 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 404

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        403 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 344

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db        343 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 284

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||
Db        283 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGAAGGTGACC 224

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db        223 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 164

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db        163 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 104

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db        103 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 44

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTT 584
              |||||||||||||  |||||||| ||||||| | ::||:|||||
Db         43 CAACCCCCTTTGGTNCCCCTCCC-TTGTATGCGASMTCKGTTTT 1


RESULT 9
BX357208/c
LOCUS       BX357208                 758 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX357208 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI022YJ18 3-PRIME, mRNA sequence.
ACCESSION   BX357208
VERSION     BX357208.2  GI:46305595
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 758)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30376125.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DI022DE09NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..758
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI022YJ18"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             73.5%;  Score 466.6;  DB 5;  Length 758;
  Best Local Similarity   88.8%;  Pred. No. 6.8e-124;
  Matches  509;  Conservative    9;  Mismatches   43;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        576 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 517

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        516 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 469

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db        468 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 409

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
```

```
                 ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        408 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 349

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db        348 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGACTGAGAGAC 289

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||
Db        288 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGAAGGTGACC 229

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db        228 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 169

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||:||||||||||||||||:||::
Db        168 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAMTAGCCAATCATCTATTSCCYS 109

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||: ::   || |||||| ||||
Db        108 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCMMSYCCCCGCGCCGGCCACGG 49

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGG 573
              |:||||||: |||||||||||||||||||||||
Db         48 CMACCCCCYCTGGGTCCCCTCCCTTTGTATGGG 16
```

```
RESULT 10
CR613169
LOCUS       CR613169                2716 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DE013YI20 of Placenta of Homo sapiens
            (human).
ACCESSION   CR613169
VERSION     CR613169.1  GI:50493976
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2716)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2716)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2716
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE013YI20"
                     /tissue_type="Placenta"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match            73.2%;  Score 464.6;  DB 3;  Length 2716;
  Best Local Similarity   90.9%;  Pred. No. 3.4e-123;
  Matches  510;  Conservative    0;  Mismatches   39;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       2168 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 2227

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       2228 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 2275

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       2276 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 2335

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db       2336 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2395

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db       2396 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGACTGAGAGAC 2455

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGAAGGTGACT 360
              |||||||||||||||||||||||  || ||||||||||||||||||||||||||||||
Db       2456 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGAAGGTGACC 2515

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db       2516 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2575

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              |||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db       2576 ACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 2635

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db       2636 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 2695

Qy        541 CAACCCCCTTTGGGTCCCCTC 561
              |||||||||||||||||||||
Db       2696 CAACCCCCTTTGGGTCCCCTC 2716
```

```
RESULT 11
BX337769/c
LOCUS       BX337769                 998 bp    mRNA    linear   EST 07-APR-2004
DEFINITION  BX337769 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YM13 3-PRIME, mRNA sequence.
ACCESSION   BX337769
VERSION     BX337769.2  GI:46272079
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 998)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 2, 2003 this sequence version replaced gi:30337641.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DI051AG07NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..998
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             71.5%;  Score 454.2;  DB 5;  Length 998;
  Best Local Similarity   89.1%;  Pred. No. 2.9e-120;
  Matches  525;  Conservative    2;  Mismatches   47;  Indels   15;  Gaps    3;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        579 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 520

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        519 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 472

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAA-GGCACCCCTCCCGA 179
              |||||||||||||||:||||| | || ||||||||||||||  || ||||||||||| ||
Db        471 CTGGACCGGCCTGCTRGCCCACGATCTGATGTTAATGACATCAAACGGCACCCCTCCTGA 412

Qy        180 GGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGAAGCAGTTAGAGCGG 239
              |||||||||| |||||||||| |||||| ||||||||||||| |||||||||||||||||
Db        411 GGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGG 352

Qy        240 TCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGA 299
              || || |||||||||||||||||||||| |||||||||||||||| ||||||||||||||
Db        351 TCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGACTGAGAGA 292

Qy        300 CAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 359
              ||||||||||||||||||||||||  || |||||||||||||||||||||||||||||||
Db        291 CAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 232

Qy        360 TGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCT 419
                ||||||||| |||||| ||||||||||||||:||||||||| ||||||||||||||||
Db        231 CACATCCACCTTTAAACACGGGGCTTGCAACTTRGCTCACACCTGACCAATCAGAGAGCT 172

Qy        420 CACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCT 479
              ||||||||||||||||||||||| | ||||||||||||||||||||||||||||||||||
Db        171 CACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCT 112

Qy        480 GAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACG 539
              ||||||||||| |||||||||| |||||||||||||||||| |  |||||||||||||||
Db        111 GAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACG 52

Qy        540 GCAACCCCCTTTGGGTCCC--CTCCCTTTGTATGGGCGCTCTGTTTTCA 586
              |||||||||||||||||||  | || ||    |||| | ||||||||||
Db         51 GCAACCCCCTTTGGGTCCCNNCNCCTTTGTANTGGGAGGTCTGTTTTCA 3


RESULT 12
BX378303/c
LOCUS       BX378303                1058 bp    mRNA    linear   EST 23-APR-2004
DEFINITION  BX378303 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI018YH07 3-PRIME, mRNA sequence.
ACCESSION   BX378303
VERSION     BX378303.2  GI:46557492
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1058)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 8, 2003 this sequence version replaced gi:30439129.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DI018CD04NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..1058
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI018YH07"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             71.4%;  Score 453.6;  DB 5;  Length 1058;
  Best Local Similarity   87.0%;  Pred. No. 4.3e-120;
  Matches  508;  Conservative    1;  Mismatches   63;  Indels   12;  Gaps    1;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        576 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 517

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db        516 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 469

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db        468 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 409

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240
              ||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||||
Db        408 GAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 349

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGACTGAGAGAC 300
              | || |||||||||||||||||||||| |||||||||||||||| |||||||||||||||
Db        348 CGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGACTGAGAGAC 289

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              |||||||||||||||||:|||||  || |||||||||||||||||||||||||||||||
Db        288 AGGACTAGCTGGATTTCYTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 229

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||||
Db        228 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 169

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
              ||||||||||||| |||||||| | |||||||||||||||||||||||||||||||||||
Db        168 ACTAAAATGCTAACTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 109

Qy        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
              |||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||||
Db        108 AGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGG 49

Qy        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTT 584
              ||||||  |   |||  |       | | || |   | |  |||
Db         48 CAACCCNNTNNTGGTNNCNNTGGCGTTTGTGTGGAGTTTTGTTT 5


RESULT 13
AG121669/c
LOCUS       AG121669                 689 bp    DNA     linear   GSS 04-NOV-2001
```

```
DEFINITION  Pan troglodytes DNA, clone: PTB-130M15.F, genomic survey sequence.
ACCESSION   AG121669
VERSION     AG121669.1  GI:16650834
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 689)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-130M15.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             71.0%;  Score 450.8;  DB 9;  Length 689;
  Best Local Similarity   88.3%;  Pred. No. 2.6e-119;
  Matches  515;  Conservative    0;  Mismatches   58;  Indels   10;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              || || ||||||||| ||||||||||| ||||||||||||||||| || ||||||||| |
Db        666 CCNTGGATCTTTAACGTCCTTGTTAAGCTTGTCTCTTCCAGAATCGAAGCTGTAAAACCA 607

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
               ||    ||||||||||||||||||  ||| ||||||||||||||||| ||| |||||||
Db        606 TAATAGTTTCTTCAAATGGAGCACCTTATGCAGTCCATGACTAAGATCTACCATGGACCC 547

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||| |||||||||||||||||| |||||||||||||| |||||||||||||| |||
Db        546 CTGGACCAGCCTGCTAGCCCATGCTCTGATGTTAATGACATCGAAGGCACCCCTCCTGAG 487

Qy        181 GAAATCTCAACTGCAC-AACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGG 239
              |||||||||||||||| ||||| |||||  | ||| |||||| |||||||||||||||||
Db        486 GAAATCTCAACTGCACAAACCCTTACTACACTCCAGTTCAGCAGGAAGCAGTTAGAGCGG 427

Qy        240 TCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGA 299
              || || ||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db        426 TCGTCGGCCAACCTCCCCAATGGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGA 367

Qy        300 CAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 359
              |||||||| ||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        366 CAGGACTATCTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 307

Qy        360 TGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC-------- 411
                ||| ||| | ||||||  ||||||||||||||||||||||||||| ||||
Db        306 CACATTCACTTTTAAACACAGGGCTTGCAACTTAGCTCACACCCGACGAATCAGGTAGTA 247
```

```
Qy        412 -AGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCAT 470
               |||| |||||||||||||||||||||||||||| |||||||||||||||||||||| ||
Db        246 AAGAGGGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATAAT 187

Qy        471 CTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAG 530
               ||||||||||||| ||||||||| |||||| ||||||||||||||| || |||||||||
Db        186 TTATTGCCTGAGAGTACAGCGGGAAGGACAATGATCGGGATATAAACTCATGCATTCGAG 127

Qy        531 CCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGG 573
              ||  ||| ||| |||| |||||||||||||||||||||   ||
Db        126 CCACCAATGGCTACCCTCTTTGGGTCCCCTCCCTTTGTTATGG 84


RESULT 14
AA781423/c
LOCUS       AA781423                 494 bp    mRNA    linear   EST 31-DEC-1998
DEFINITION  aj26c03.s1 Soares_testis_NHT Homo sapiens cDNA clone 1391428 3'
            similar to contains PTR7.t1 PTR7 repetitive element ;, mRNA
            sequence.
ACCESSION   AA781423
VERSION     AA781423.1  GI:2840754
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1645   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 475.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="1391428"
                     /sex="male"
                     /lab_host="DH10B"
                     /clone_lib="Soares_testis_NHT"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

  Query Match             68.8%;  Score 437;  DB 1;  Length 494;
  Best Local Similarity   92.7%;  Pred. No. 2.4e-115;
  Matches  458;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy        131 CTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAA 190
              ||||||||||| | || ||||||||||||||  ||||||||||||| ||||||||||||
Db        494 CTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAATCTCAG 435
```

```
Qy       191 CTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAA 250
             |||||||||| |||||| ||||||||||||| ||||||||||||||||||| || |||||
Db       434 CTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTCGGCCAA 375

Qy       251 CCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCT 310
             ||||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||
Db       374 CCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGACAGGACTAGCT 315

Qy       311 GGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCT 370
             |||||||||||||  || |||||||||||||||||||||||||||||||  |||||||||
Db       314 GGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACATCCACCT 255

Qy       371 CTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAAATGC 430
              |||||| |||||||||||||||||||||||| |||||||||||||||||||||||||||
Db       254 TTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATGC 195

Qy       431 TAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGC 490
             ||| |||||||| | |||||||||||||||||||||||||||||||||||||||||||||
Db       194 TAACTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGC 135

Qy       491 GGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCTT 550
              |||||||||| |||||||||||||||||| |  |||||||||||||||||||| |||||
Db       134 AGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGGCAACNCCCTT 75

Qy       551 TGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCTT 610
             |||||||||||||||||| |||| ||||||||||||  ||||||||||||||||||||||
Db        74 TGGGTCCCCTCCCTTTGTGTGGGAGCTCTGTTTTCATGCTATTTCACTCTATTAAATCTT 15

Qy       611 GCAACTGAAAAAAA 624
             ||||||| ||||||
Db        14 GCAACTGCAAAAAA 1
```

```
RESULT 15
BX439636/c
LOCUS       BX439636                1019 bp    mRNA    linear   EST 04-MAY-2004
DEFINITION  BX439636 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012YJ24
            3-PRIME, mRNA sequence.
ACCESSION   BX439636
VERSION     BX439636.2  GI:47000005
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1019)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 15, 2003 this sequence version replaced gi:30771765.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DE012DE12NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..1019
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YJ24"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             68.5%;  Score 435;  DB 5;  Length 1019;
  Best Local Similarity   83.4%;  Pred. No. 1.1e-114;
  Matches  494;  Conservative   19;  Mismatches   66;  Indels   13;  Gaps    2;
```

```
Qy         1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
             |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db       579 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 520

Qy        61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                         |||||||||| | ||||| |||||| |||||||||| ||||  |||||
Db       519 ------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCC 472

Qy       121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
             ||||||||||||||||||||| | || ||||||||||||||  ||||||||||||| |||
Db       471 CTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAG 412

Qy       181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
             ||||||||| |||:||||||  ||||| ||||||||||||| ||||||||||||||||||
Db       411 GAAATCTCAGCTGMACAACCTATACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 352

Qy       241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
             | || |||||||||||| ||||||||| |||||||||||||||| |||||||||||||||
Db       351 CGTCGGCCAACCTCCCCCACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 292

Qy       301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
             ||||||||||||||: ||:||||  || ||||||||||||||||||||||||||||||| 
Db       291 AGGACTAGCTGGATYCCCYAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC 232

Qy       361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
              ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||| 
Db       231 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTK 172

Qy       421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
             ||||||||||||| |||||||| | ||||||:::: ||||: ::|| || |||||||||
Db       171 ACTAAAATGCTAACTAGGCAAAKACAGGAGGTWWWKTAATAKNVRATTATTTATTGCCTG 112

Qy       481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540
             |||||||||| ||||||||:| :|||||||||||||:||| |  |||||||||||||| |
Db       111 AGAGCACAGCAGGAGGGACWATKATCGGGATATAAAYCCAAGTCTTCGAGCCGGCAACAG 52

Qy       541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTAT 592
              ||||:::|||   ||:|: |  ||||| |||| : | | |||    | | |
Db        51 AAACCYYYTTT-TTTCYCYCCTTTTTGTNTGGGNDTTTTTTTTNNTTTATTT 1
```

```
RESULT 16
AW971553/c
LOCUS       AW971553                 522 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST383642 MAGE resequences, MAGL Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW971553
VERSION     AW971553.1  GI:8161399
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C.,
            Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 292
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1..522
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGL"
                     /note="Vector: pBluescriptSKm"
ORIGIN

  Query Match             68.1%;  Score 432.4;  DB 2;  Length 522;
  Best Local Similarity   91.8%;  Pred. No. 5.2e-114;
  Matches  479;  Conservative    0;  Mismatches   41;  Indels    2;  Gaps    2;

Qy         97 ATGACTAAGATCCACCGTGGACCCCT-GGACCGGCCTGCTAGCCCATGCTCCGATGTTAA 155
              | |  |||||||  ||| ||||||||   || ||| |||||||||||| || | ||||||
Db        522 AGGTTTAAGATCTTCCGCGGACCCCTAAAACTGGCTTGCTAGCCCATGTTCTGGTGTTAA 463

Qy        156 TGACATTGAAGGCACCCCTCCCGAGGAAA-TCTCAACTGCACAACCCCTACTATGCCCCA 214
              |||||| |||| ||| ||||| ||||||| ||||||||||||||||| || |||||||||
Db        462 TGACATCGAAGTCACTCCTCCTGAGGAAATTCTCAACTGCACAACCCTTATTATGCCCCA 403

Qy        215 ATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCACTTGGGTTTT 274
               |||||| |||||||||||||| ||||||||||||||||| |||||||||||||||||||
Db        402 GTTCAGCAGGAAGCAGTTAGAGTGGTCATCAGCCAACCTCACCAACAGCACTTGGGTTTT 343

Qy        275 CCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAAT 334
              ||||||||||| |||||||||||||||||||||||||||||||||||||| |  ||||||
Db        342 CCTGTTGAGAGAGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGATTAAGAAT 283

Qy        335 CCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAG 394
              |||||| |||||||||||||||||| |||||||||| |||||| ||||||||||||||||
Db        282 CCCTAAACCTAGCTGGGAAGGTGACCGCATCCACCTTTAAACACGGGGCTTGCAACTTAG 223

Qy        395 CTCACACCCGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAA 454
              ||||||||| |||||||||||||||||||||||||||||||||||||||| |||||||||
Db        222 CTCACACCCAACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAA 163

Qy        455 AGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATA 514
              ||||||||||||||||||||||||||||||||||| ||||||||||||||| |  |||||
Db        162 AGAAATAGCCAATCATCTATTGCCTGAGAGCACAGTGGGAGGGACAAGGATTGCAATATA 103

Qy        515 AACCCAGGCATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGC 574
              |||||||||||||||||| |||||||||||| |||||||||||| ||||||||||||||
Db        102 AACCCAGGCATTCGAGCCAGCAACGGCAACCGCCTTTGGGTCCCTTCCCTTTGTATGGGA 43

Qy        575 GCTCTGTTTTCACTCTATTTCACTCTATTAAATCTTGCAACT 616
              ||||||||||||||||||||||||||||||||||||||||||
Db         42 GCTCTGTTTTCACTCTATTTCACTCTATTAAATCTTGCAACT 1


RESULT 17
CN645411
LOCUS       CN645411                870 bp    mRNA    linear   EST 13-MAY-2004
DEFINITION  ILLUMIGEN_MCQ_23968 Katze_MMSP Macaca mulatta cDNA clone
            IBIUW:10227 5' similar to Bases 165 to 770 highly similar to human
            Unigene Hs.349001, mRNA sequence.
ACCESSION   CN645411
VERSION     CN645411.1  GI:47158854
KEYWORDS    EST.
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 870)
  AUTHORS   Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
  TITLE     Large-scale Rhesus Macaque cDNA Sequencing
  JOURNAL   Unpublished (2003)
COMMENT     Contact: C. Magness
            Illumigen Biosciences Inc.
            2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
            Tel: 2063780400
            Fax: 2063780408
            Email: cmagness@illumigen.com
            Sequenced on 2004.02.24. 676 Q20 bases.
            PCR PRimers
            FORWARD: CCCTCACTAAAGGGAACAAAA
            BACKWARD: CACTATAGGGCGAATTGGGTA
            Insert Length: 870   Std Error: 0.00
            Plate: CL000135  row: H  column: 02
            Seq primer: CCCTCACTAAAGGGAACAAAA
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..870
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /strain="Indian"
                     /db_xref="taxon:9544"
                     /clone="IBIUW:10227"
                     /sex="male"
                     /cell_type="mononuclear lymphocyte"
                     /dev_stage="adult"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Katze_MMSP"
                     /note="Organ: spleen; Vector: Uni-ZAP XR; Site_1: EcoR I;
                     Site_2: Xho I; Created from Stratagene ZAP-cDNA Synthesis
                     kit (catalog #200400)  and ZAP-CDNA Gigapack III Gold
                     Cloning Kit (Catalog #200450)"
ORIGIN

  Query Match             67.7%;  Score 430.2;  DB 7;  Length 870;
  Best Local Similarity   83.8%;  Pred. No. 2.5e-113;
  Matches  527;  Conservative    0;  Mismatches   88;  Indels   14;  Gaps    3;

Qy         21 TGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATGGA 80
              | |||| ||||| |||||||||||| || |||||||||||||||| ||||||||||||||
Db        166 TTTTAAATTTGTTTCTTCCAGAATCGAAGCTGTAAAACTACAAATGGTTCTTCAAATGGA 225

Qy         81 GCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAGCCC 140
              ||  | |||| ||||||||||||||||| || | ||||||||||||| ||||||||  ||
Db        226 GCCTCCGATGCAGTCCATGACTAAGATCTACTGCGGACCCCTGGACCAGCCTGCTAATCC 285

Qy        141 ATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTGCACAACC 200
              |||||| ||||||  |||||| |||||||||||||||||||||||||||||||||  |||
Db        286 ATGCTCTGATGTTGGTGACATCGAAGGCACCCCTCCCGAGGAAATCTCAACTGCATGACC 345

Qy        201 CCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAAC 260
              ||||||| ||| || |||||| ||||||||||||||| ||||||||||||||||| ||||
Db        346 CCTACTACGCCTCAGTTCAGCAGGAAGCAGTTAGAGCAGTCATCAGCCAACCTCCTCAAC 405

Qy        261 AGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTA 320
              ||||||||| |||||||||||  |||||||||||||||||||||||||||||||||||||
Db        406 AGCACTTGGATTTTCCTGTTGTAAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTA 465

Qy        321 GGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACATGG 380
               |||||| ||||||||||||||||||||||||||||||||||||||| || |||||| ||
Db        466 AGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCATCTTTAAACACGG 525

Qy        381 GGCTTGCAACTTAGCTCACACCCGACCAATC---------AGAGAGCTCACTAAAATGCT 431
              ||| ||||| ||||||||||||||| |||||         |||||||||||||||||| |
Db        526 GGCCTGCAATTTAGCTCACACCCGATCAATCAGGTAGTAAAGAGAGCTCACTAAAATGTT 585

Qy        432 AATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATC--TATTGCCTGAGAGCACAG 489
              ||||| ||||||| ||||||||||||||||| |||||| |  ||| ||||||||||||||
Db        586 AATTAAGCAAAAACAGGAGGTAAAGAAATAGTCAATCACCCTTATCGCCTGAGAGCACAG 645

Qy        490 CGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCT 549
                ||| |||||| ||| ||||||||||||||||||||| || | |||| |||||||||||
Db        646 GAGGAAGGACAATGATAGGGATATAAACCCAGGCATTCAAGTCAGCAATGGCAACCCCCT 705

Qy        550 TT---GGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAA 606
              ||   |   |||  ||| || |||||| ||||||  ||  |||||||  |   |   |
Db        706 TTTTGGTCCCCCTCCCCATTTTATGGGAGCTCTGGGTTTCCTCTATTAAATCTTGCAATT 765
```

```
Qy          607 TCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
                 ||   ||   | | |  |||||| ||||
Db          766 GCTCAAAAAAAACATATTAAAAAATAAAA 794


RESULT 18
BC026287
LOCUS       BC026287                1500 bp    mRNA    linear   HTC 04-MAR-2003
DEFINITION  Homo sapiens, clone IMAGE:4770655, mRNA.
ACCESSION   BC026287
VERSION     BC026287.1  GI:22382161
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1500)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 37 Row: m Column: 12
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1..1500
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4770655"
                     /tissue_type="Placenta"
                     /clone_lib="NIH_MGC_79"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
ORIGIN

  Query Match             66.8%;  Score 424;  DB 3;  Length 1500;
  Best Local Similarity   85.7%;  Pred. No. 1.8e-111;
  Matches  557;  Conservative    0;  Mismatches   75;  Indels   18;  Gaps    7;

Qy            1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
                |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |
Db          849 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACAA 908

Qy           61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                || || ||||||||||||||| ||||||| |||||||||||||||||| |||| ||| ||
Db          909 CACATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACCGCGGATCC 968

Qy          121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
                ||||||| ||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db          969 CTGGACCAGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACTCCTCCCGAG 1028

Qy          181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
                |||||||||| |||||||||||||||||||||||||| ||| |||||||||| |||| ||
Db         1029 GAAATCTCAAATGCACAACCCCTACTATGCCCCAATTAAGCAGGAAGCAGTTGGAGCAGT 1088

Qy          241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                |  | |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         1089 CGACGGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAGAGAC 1148

Qy          301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCT-GGGAAGGTGAC 359
                |||||||||||||||||||||||| || || |||  |||||| ||| | |||||||||||
Db         1149 AGGACTAGCTGGATTTCCTAGGCCGACTAACAATTTCTAAGCTTAGGTGGGGAAGGTGAC 1208

Qy          360 TGCATCCACCTCTAAACATGGGGC--TTGCAACTTAGCTCACACCCGACCAATC------ 411
                |||| | |||| ||||||  || |  ||| |||| |||||||| | | ||||||
Db         1209 TGCACCTACCTTTAAACACAGGACTTTTGTAACTCAGCTCACATCTGGCCAATCAGGTAG 1268

Qy          412 ---AGAGAGCTCACTAAAATGCTAATTAGGC-AAAAATAGGAGGTAAAGAAATAGCCAAT 467
                   |||||||||| |||||| | |||||| | ||||| ||||||||||||||||   |||
Db         1269 TAAAGAGAGCTCATTAAAATACCAATTAGTCTAAAAACAGGAGGTAAAGAAATA---AAT 1325

Qy          468 CATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAA-CCCAGGCATT 526
                ||||| | |||||| |||||||| || |||| ||| ||| ||||||||| ||||||||||
Db         1326 CATCTCTCGCCTGACAGCACAGAGGAAGGGGCAATGATTGGGATATAAACCCCAGGCATT 1385

Qy          527 CGAGCCGGCA-ACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTC 585
                ||||| || |   |||||||||||||||||||||||||| |||||||| ||||||||||||
Db         1386 CGAGCTGGGAGTGGGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTC 1445

Qy          586 ACTCTATTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
                ||||||||  |   |   |   | |  ||   ||||||||||||||||||
Db         1446 ACTCTATTCAATCTTGCAACTGCATAAAAAAAAAAAAAAAAAAAAAAAAA 1495


RESULT 19
AG033781/c
LOCUS       AG033781                 653 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-008021.F, genomic survey sequence.
ACCESSION   AG033781
VERSION     AG033781.1  GI:16560654
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 653)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector     : pKS145
                R.Site 1   : SacI
                R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1..653
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-008021.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
```

```
  Query Match             66.6%;  Score 423;  DB 9;  Length 653;
  Best Local Similarity   88.0%;  Pred. No. 2.9e-111;
  Matches  511;  Conservative    0;  Mismatches   52;  Indels   18;  Gaps    4;

Qy            2 CCTGTATCTTTAACCTCC-TTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
                ||||| ||||||||| || ||||||||| ||||||| ||||||||| | ||||||| |||
Db          639 CCTGTTTCTTTAACCCCCGTTGTTAAGTGTGTCTCTNCCAGAATCANAGCTGTAAAGCTA 580

Qy           61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                ||||| |||||||||||||||| ||| || |                   || |||||||
Db          579 CAAATGGTTCTTCAAATGGAGCCCCAAATTG---------------CAGTCCATGGACCC 535

Qy          121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
                |||||||||||||||||||||||||| |||||||||||||| ||||||  |||||| |||
Db          534 CTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCGAAGGC-GCCCTCCTGAG 476

Qy          181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
                 ||||||||||||||||||||||||||  ||||| |||||| |||||||||||||| |||
Db          475 AAAATCTCAACTGCACAACCCCTACTACACCCCAGTTCAGCAGGAAGCAGTTAGAGTGGT 416

Qy          241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
                | |  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          415 CGTTGGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 356

Qy          301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
                |||||||||||||||||||| |||||| ||||||||||||||||| |||||||||||||
Db          355 AGGACTAGCTGGATTTCCTAAGCCAACTAAGAATCCCTAAGCCTACCTGGGAAGGTGACC 296

Qy          361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCA-GAGAGCT 419
                 ||||||||| |||||||||||||||||||||||||||||||| |||| | | |||||||
Db          295 ACATCCACCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCAACCAGTAAGGAGAGCT 236

Qy          420 CACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCT 479
                ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          235 CACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCT 176

Qy          480 GAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACG 539
                ||||||||||| | || ||||| ||||| |||||||||||| ||||||||||| |||| |
Db          175 GAGAGCACAGCAGAAGAGACAATGATCGAGATATAAACCCAAGCATTCGAGCCAGCAATG 116

Qy          540 GCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTG 580
                || |||  |||||||||||||||||||||||||| |||| |
Db          115 GCTACCTTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCCG 75


RESULT 20
AI598135/c
LOCUS       AI598135                 490 bp    mRNA    linear   EST 12-MAY-1999
DEFINITION  tn14a10.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2167578 3'
            similar to contains PTR5.t1 PTR5 repetitive element ;, mRNA
            sequence.
ACCESSION   AI598135
VERSION     AI598135.1  GI:4607183
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1407   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 455
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..490
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2167578"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Brn25"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
ORIGIN

  Query Match             65.5%;  Score 415.8;  DB 1;  Length 490;
  Best Local Similarity   92.9%;  Pred. No. 3.4e-109;
  Matches  435;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          126 CCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAAT 185
                || |||||||||||| || |||||||||||||||||  ||||||||||||| ||||||||
Db          490 CCAGCCTGCTAGCCCTTGTTCCGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAAT 431

Qy          186 CTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCA 245
                |||| ||||||||||||||| |||||||| |||||| |||||||| |||||||||| |||
Db          430 CTCANCTGCACAACCCCTACCATGCCCCATTTCAGCAGGAAGCAGGTAGAGCGGTCGTCA 371

Qy          246 GCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGAC 305
                | ||||||||||| |||||||||||||||||||||||||||||| |||||||||||||||
Db          370 GTCAACCTCCCCAGCAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAGAGACAGGAC 311

Qy          306 TAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATC 365
                ||||||||||||||||||| || ||||||||||||||||||||||||||||||| || ||
Db          310 TAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCGCGTC 251

Qy          366 CACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAA 425
                ||||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db          250 CACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAA 191

Qy          426 AATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGC 485
                ||||||||||| ||||||| |||||||||||||||||||| |||||||| ||||||||||
Db          190 AATGCTAATTAAGCAAAAACAGGAGGTAAAGAAATAGCCAGTCATCTATCGCCTGAGAGC 131

Qy          486 ACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACC 545
                |||||  ||||||||| |||||||||||||||||||||||||||||| |||||||| |||
Db          130 ACAGCAAGAGGGACAATGATCGGGATATAAACCCAGGCATTCGAGCCAGCAACGGCTACC 71

Qy          546 CCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATT 593
                | ||||||||||||||||||||||||||| || ||||||||||||||||
Db           70 CTCTTTGGGTCCCCTCCCTTTGTATGGGAGCCCTGTTTTCACTCTATT 23


RESULT 21
BE732673
LOCUS       BE732673                 440 bp    mRNA    linear   EST 15-SEP-2000
DEFINITION  601571305F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3925728 5',
            mRNA sequence.
ACCESSION   BE732673
VERSION     BE732673.1  GI:10146665
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 440)
 AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
 TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL  Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
          Plate: LLCM752  row: j  column: 01
          High quality sequence stop: 440.
FEATURES           Location/Qualifiers
    source         1..440
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:3925728"
                   /tissue_type="choriocarcinoma"
                   /lab_host="DH10B (phage-resistant)"
                   /clone_lib="NIH_MGC_21"
                   /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                   Site_2: EcoRI; cDNA made by oligo-dT priming.
                   Directionally cloned into EcoRI/XhoI sites using the
                   following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                   for average insert size 1.8kb.  Library constructed by
                   Ling Hong in the laboratory of Gerald M. Rubin (University
                   of California, Berkeley) using  ZAP-cDNA synthesis kit
                   (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

  Query Match             62.6%;  Score 397.4;  DB 2;  Length 440;
  Best Local Similarity   94.1%;  Pred. No. 7.1e-104;
  Matches  413;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy        179 AGGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCG 238
              ||||||||||| |||||||||| |||||| ||||||||||||| ||||||||||||||||
Db          1 AGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCG 60

Qy        239 GTCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAG 298
              ||| || |||||||||||||||||||||| |||||||||||||||| |||||||||||||
Db         61 GTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAG 120

Qy        299 ACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGA 358
              |||||||||||||||||||||||||  || ||||||||||||||||||||||||||||||
Db        121 ACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGA 180

Qy        359 CTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGC 418
              |  ||||||||| |||||| |||||||||||||||||||||||| |||||||||||||||
Db        181 CCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGC 240

Qy        419 TCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCC 478
              |||||||||||||||||||||||| | |||||||||||||||||||||||||||||||||
Db        241 TCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCC 300

Qy        479 TGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAAC 538
              |||||||||||| |||||||||| |||||||||||||||||| |  ||||||||||||||
Db        301 TGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAAC 360

Qy        539 GGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACT 598
              ||||||||||||||||||||||||||||||||||| ||||||||||||  ||||||||||
Db        361 GGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACT 420

Qy        599 CTATTAAATCTTGCAACTG 617
              |||||||||||||||||||
Db        421 CTATTAAATCTTGCAACTG 439
```

```
RESULT 22
BI963185/c
LOCUS       BI963185                 543 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  ie57c10.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA clone IMAGE:5670930 3', mRNA sequence.
ACCESSION   BI963185
VERSION     BI963185.1  GI:16337590
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu) This sequence now available from the IMAGE
            consortium, for clone orders contact: info@image.llnl.gov
            High quality sequence stop: 451.
FEATURES             Location/Qualifiers
     source          1..543
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5670930"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit  (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by  column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified  once on solid support and plasmid DNA from
                     library was prepared. The  library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library  inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized)  plasmids were isolated
                     by hydroxyapatite chromatography and used to make  this
                     library."
ORIGIN

  Query Match             62.4%;  Score 396;  DB 4;  Length 543;
  Best Local Similarity   88.3%;  Pred. No. 1.9e-103;
  Matches  467;  Conservative    0;  Mismatches   51;  Indels   11;  Gaps    3;

Qy         74 AAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCCTG 133
              ||||||||| ||||||| ||||||||||||||||| || | ||||| |||||||||||||
Db        543 AAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACTGCGGACCTCTGGACCGGCCTG 484

Qy        134 CTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTG 193
              |||||||| ||||| |  ||||||||||  ||||||||||||||||||||||||||||||
Db        483 CTAGCCCA-GCTCCCACATTAATGACATCAAAGGCACCCCTCCCGAGGAAATCTCAACTG 425
```

```
Qy        194 CACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCT 253
              ||  || ||||||| ||||||||||||| |||||||||||||| |||| |  ||||||||
Db        424 CATGACACCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGTCGTTGGCCAACCT 365

Qy        254 CCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGA 313
              ||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db        364 CCCCAACAGCACTTGGGTTTT-CTGTTGAGATGGGGGACTGAGAGACAGGACTAGCTGGA 306

Qy        314 TTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTA 373
              |||||||||||||| |||||||||||||||||||||||||||||||| || |||||| ||
Db        305 TTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTACACCCACCTTTA 246

Qy        374 AACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT---------CAGAGAGCTCACTA 424
              | ||   |||||||||||||||||||||||||||| |          ||||||||  |||
Db        245 ACCACTAGGCTTGCAACTTAGCTCACACCCGACCANTAGGGTAGTAAAGAGAGCTTGCTA 186

Qy        425 AAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAG 484
              |||||||||||||||||||| |||||||||||||||||||| |||||||| |||||| ||
Db        185 AAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAGTCATCTATCGCCTGACAG 126

Qy        485 CACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAAC 544
              ||||  ||| ||||||| |||| |||||||||| ||||||||| |||| |||| ||| ||
Db        125 CACAAGGGGCGGGACAATGATCAGGATATAAACTCAGGCATTCAAGCCAGCAATGGCTAC 66

Qy        545 CCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATT 593
              || |||||||||||||||| || |||||| |||||||||||||||||||
Db         65 CCACTTTGGGTCCCCTCCCATTTTATGGGAGCTCTGTTTTCACTCTATT 17


RESULT 23
BX380176/c
LOCUS       BX380176                 966 bp    mRNA    linear   EST 28-APR-2004
DEFINITION  BX380176 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI044YK06 3-PRIME, mRNA sequence.
ACCESSION   BX380176
VERSION     BX380176.2  GI:46834414
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 966)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 8, 2003 this sequence version replaced gi:30460243.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DI044BF03NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..966
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI044YK06"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
```

```
  Query Match             62.2%;  Score 395.2;  DB 5;  Length 966;
  Best Local Similarity   79.6%;  Pred. No. 3.7e-103;
  Matches  465;  Conservative   22;  Mismatches   84;  Indels   13;  Gaps    2;

Qy          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60
              |||||||||||||||||||||||||| |||||||||||||||||| || |||||||||||
Db        574 CCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAACTA 515

Qy         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
                          |||||||||| | :|||| |||||| ||||| |:|| |||| || |||
Db        514 ------------CAAATGGAGCCCARGATGCAGTCCARGACTAGGDTCTACCGCGGCCCC 467

Qy        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
              ||||||||||||||| ||||  |:|| ||||||||||||||  ||||||||||||| | |
Db        466 CTGGACCGGCCTGCTSGCCCCCGMTCTGATGTTAATGACATCAAAGGCACCCCTCCTGGG 407

Qy        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
              | ||||||| ||||||: || |||||| |||||||||| ||:||||||||||:| |||||
Db        406 GCAATCTCAGCTGCACMCCCTCTACTACGCCCCAATTCGGCRGGAAGCAGTTRGGGCGGT 347

Qy        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
              | || |||||||||||| ||:||||||:|||||||||||||||| |||||||||||||||
Db        346 CGTCGGCCAACCTCCCCGACRGCACTTRGGTTTTCCTGTTGAGATGGGGGACTGAGAGAC 287

Qy        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
              :||||| ||||||||||||||||  || |||||||||||:||||:||||||||||||||
Db        286 MGGACTGGCTGGATTTCCTAGGCTGACTAAGAATCCCTARGCCTRGCTGGGAAGGTGACC 227

Qy        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
               ||||||||| |||||: ||||||||| |||| |||||| || | ||||||:| | ||||
Db        226 ACATCCACCTTTAAACMCGGGGCTTGCCACTTSGCTCACCCCTGCCCAATCRGGGGGCTC 167

Qy        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGC-CAATCATCTATTGCCT 479
              ||||||||||||:|: |||       :: ::    |   | || |: :|:||||||||||
Db        166 ACTAAAATGCTAMTYCGGCCCCGSCSSSSSVCSSSGCCCTCGCGCMSYCMTCTATTGCCT 107

Qy        480 GAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACG 539
              ||||||||||| |||||||||| |||||||||||||||||| |  |||| ||||||  ||
Db        106 GAGAGCACAGCMGGAGGGACAATGATCGGGATATAAACCCAMGTCTTCGCGCCGGCCCCG 47

Qy        540 GCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTT 583
              ||  |||||||||||||||||||||||      | |  || | |
Db         46 GCCCCCCCCTTTGGGTCCCCTCCCTTTCGTNNTGGGAGCTCTGT 3


RESULT 24
AG058970/c
LOCUS       AG058970                 651 bp    DNA     linear   GSS 02-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-046A08.R, genomic survey sequence.
ACCESSION   AG058970
VERSION     AG058970.1  GI:16596431
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 651)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
                Sequencing: M13Rev
            LIBRARY
                Vector     : pKS145
                R.Site 1   : SacI
                R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-046A08.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             62.2%;  Score 394.8;  DB 9;  Length 651;
  Best Local Similarity   87.6%;  Pred. No. 4.4e-103;
  Matches  503;  Conservative    0;  Mismatches   57;  Indels   14;  Gaps    6;

Qy           20 TTGTTAAGTTTGTCTCTTCCA-GAATCAAAACTGTAAAACTACAAATTGTTCTTCAAATG 78
                ||| || |||||||||| ||| |||||||| |||||||||||||||| ||||||||||||
Db          613 TTGATAGGTTTGTCTCTCCCAAGAATCAAAGCTGTAAAACTACAAATGGTTCTTCAAATG 554

Qy           79 GAGCACCAGATGGAGTCCATGAC-TAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAG 137
                |||||||||||  ||| |||| | |||| ||  ||||| |||||||||| |||| |||||
Db          553 GAGCACCAGATACAGTGCATGTCTTAAGTTCTCCCGTGAACCCCTGGACTGGCCAGCTAG 494

Qy          138 CCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTGCACA 197
                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          493 CCCATGCTCTGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAAATCTCAACTGCACA 434

Qy          198 ACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCC 257
                |||||| |||  |||||||||||| ||||||||||||||| ||| |||||||||||||||
Db          433 ACCCCTGCTACACCCCAATTCAGCAGGAAGCAGTTAGAGCAGTCGTCAGCCAACCTCCCC 374

Qy          258 AACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTC 317
                |||||||||||||||||||||||||||| |||||||||||||||| ||||||||||||||
Db          373 AACAGCACTTGGGTTTTCCTGTTGAGAGTGGGGACTGAGAGACAGAACTAGCTGGATTTC 314

Qy          318 CTAGGCCAACGAAGAATCCCTAAGCCTAGCT-GGGAAGGTGACTGCATCCACCTCTAAAC 376
                ||||||  || |||||| ||||||||||||| |||||||||| |||| |||||| |||||
Db          313 CTAGGCTGACTAAGAATTCCTAAGCCTAGCTGGGGAAGGTGATTGCACCCACCTTTAAAC 254

Qy          377 ATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAG---------AGAGCTCACTAAAA 427
                |||||||||| |||| ||||||||||| || ||||||         || |||||||||||
Db          253 ATGGGGCTTGTAACTCAGCTCACACCCAACAAATCAGGTAGTAAAAAGGGCTCACTAAAA 194

Qy          428 TGCTAATTAGGC-AAAAATAGGAGGTAAAGAAATAG-CCAATCATCTATTGCCTGAGAGC 485
                | | |||||||| ||||  ||||| ||||||||||| | |||||| |||  ||||||| |
Db          193 TACAAATTAGGCTAAAAGCAGGAGATAAAGAAATAGTCAAATCATATATCACCTGAGACC 134

Qy          486 ACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACC 545
                |||| ||||||||||| | ||||| ||||||||||||||||||||| || |  |||||||
Db          133 ACAGTGGGAGGGACAATGGTCGGGGTATAAACCCAGGCATTCGAGCAGGGAGTGGCAACC 74

Qy          546 CCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCT 579
                | |||||||||||||||| ||||||||| |||||
Db           73 CGCTTTGGGTCCCCTCCCATTGTATGGGAGCTCT 40


RESULT 25
AG126669/c
LOCUS       AG126669                 701 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-137E19.F, genomic survey sequence.
ACCESSION   AG126669
VERSION     AG126669.1  GI:16655834
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
```

```
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 701)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector     : pKS145
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-137E19.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             61.9%;  Score 393.2;  DB 9;  Length 701;
  Best Local Similarity   88.1%;  Pred. No. 1.3e-102;
  Matches  465;  Conservative    0;  Mismatches   53;  Indels   10;  Gaps    3;

Qy           79 GAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCCTGCTAGC 138
                || | | ||||  |||||||||||||   | || ||| |||||||||| |||||||||||
Db          651 GACCCCGAGATCCAGTCCATGACTAAAGACTACTGTGAACCCCTGGACTGGCCTGCTAGC 592

Qy          139 CCATGCTCCGATGTTAATGACATTGAAGGCACCCCT-CCCGAGGAAATCTCAACTGCACA 197
                ||||||| | |||||||||||||||||||||||||| |||||||||||||||||||||||
Db          591 CCATGCTGCAATGTTAATGACATTGAAGGCACCCCTGCCCGAGGAAATCTCAACTGCACA 532

Qy          198 ACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCC 257
                |||||||||||||||| ||||||| ||| ||||||| || |||| |  ||||| ||||||
Db          531 ACCCCTACTATGCCCCGATTCAGCAGGAGGCAGTTAAAGTGGTCGTTGGCCAAACTCCCC 472

Qy          258 AACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTC 317
                ||||||| |||| ||||||||||||||||||  |||||||||||||||||||||||||||
Db          471 AACAGCAGTTGGTTTTTCCTGTTGAGAGGGGATACTGAGAGACAGGACTAGCTGGATTTC 412

Qy          318 CTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACA 377
                ||||||  || ||||||||||||| |||||||||||||||||  |||||||||  |||||
Db          411 CTAGGCTGACTAAGAATCCCTAAGACTAGCTGGGAAGGTGACCACATCCACCTTGAAACA 352

Qy          378 TGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------AGAGAGCTCACTAAAATG 429
                 |||||||||||||||||||||| ||||||||||        ||| || | |||||||||
Db          351 CGGGGCTTGCAACTTAGCTCACATCCGACCAATCAGGTAATAAGAAAGATAACTAAAATG 292

Qy          430 CTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAG 489
                ||||||||||||||| |||||||||||||||||||||||||||||  ||||||| |||||
Db          291 CTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATCACCTGAGATCACAG 232

Qy          490 CGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCT 549
                | |||||||||| |||||||||||||||||||||||||||||| ||||| || |||| ||
Db          231 CAGGAGGGACAATGATCGGGATATAAACCCAGGCATTCGAGCCAGCAACAGCTACCCTCT 172

Qy          550 TTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCAC 597
                || ||||||||||||||||||||| |||||||||||||||||||  ||
Db          171 TT-GGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTAAAC 125
```

```
RESULT 26
AG066901/c
LOCUS       AG066901                 609 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-056L15.F, genomic survey sequence.
ACCESSION   AG066901
VERSION     AG066901.1  GI:16618703
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 609)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector     : pKS145
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1. .609
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-056L15.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             61.9%;  Score 393;  DB 9;  Length 609;
  Best Local Similarity   87.7%;  Pred. No. 1.4e-102;
  Matches  478;  Conservative    0;  Mismatches   55;  Indels   12;  Gaps    4;

Qy          47 AAACTGTAAAACTACAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGA 106
               || ||||| |||||||||  ||||||||||||||| ||| ||| ||||||||||||||||
Db         606 AAGCTGTAGAACTACAAATCATTCTTCAAATGGAGCCCCAAATGCAGTCCATGACTAAGA 547

Qy         107 TCCACCGTGGACCCCTGGACCGGCCTGCTAGCCCATGCTCC-GATGTTAATGACATTGAA 165
               || || |  ||||| |||||| |||||||||| ||||||||  ||||||||| | |||||
Db         546 TCTACTGCAGACCCTTGGACCAGCCTGCTAGCTCATGCTCCAAATGTTAATGGCCTTGAA 487

Qy         166 GGCACC-CCTCCCGAGGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGG 224
                || ||  ||||| |||||||||||||||||| ||||||||||  |||||||||||| ||
Db         486 AGCCCCTTCTCCCAAGGAAATCTCAACTGCACCACCCCTACTACACCCCAATTCAGCAGG 427

Qy         225 AAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGA 284
               ||||||||||||| |||||| |||||||||||||||||||||||||||||| ||||||||
Db         426 AAGCAGTTAGAGCAGTCATCGGCCAACCTCCCCAACAGCACTTGGGTTTTCATGTTGAGA 367

Qy         285 GGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCT 344
                |||||||| |||||||||||||||||||||||||| ||  || ||||||||||||||||
Db         366 TGGGGGACTTAGAGACAGGACTAGCTGGATTTCCTA-GCTGACTAAGAATCCCTAAGCCT 308

Qy         345 AGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCG 404
               | |||||||||||||  | ||||||| |||||| |||||||||||||||||||||||  |
Db         307 ACCTGGGAAGGTGACCACTTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACATG 248

Qy         405 ACCAATCA---------GAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAA 455
               ||||||||         ||||||||||||||||||||||||||||| || ||||||||||
Db         247 ACCAATCAGATAGTAAGGAGAGCTCACTAAAATGCTAATTAGGCAACAACAGGAGGTAAA 188

Qy         456 GAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGAGGGACAAGGATCGGGATATAA 515
               |||||||||||||||||  | ||||| ||||||| |||||||||| ||| ||||||||||
Db         187 GAAATAGCCAATCATCTGCTACCTGAAAGCACAGTGGAGGGACAATGATTGGGATATAA 128

Qy         516 ACCCAGGCATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCG 575
               ||||||||||| ||||| |||| ||||||||||||||||||||||||||||||||||| |
Db         127 ACCCAGGCATTTGAGCCAGCAATGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAG 68

Qy         576 CTCTG 580
               ||| |
Db          67 CTCCG 63


RESULT 27
AA837267/c
LOCUS       AA837267                 443 bp    mRNA    linear   EST 31-MAR-1998
DEFINITION  od26b10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1369051
            similar to contains PTR7.t1 PTR7 repetitive element ;, mRNA
            sequence.
ACCESSION   AA837267
VERSION     AA837267.1  GI:2912466
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1375   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1369051"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GCB1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN
```

```
  Query Match             61.8%;  Score 392.6;  DB 1;  Length 443;
  Best Local Similarity   94.1%;  Pred. No. 1.8e-102;
  Matches  418;  Conservative    0;  Mismatches   25;  Indels    1;  Gaps    1;

Qy        184 ATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCAT 243
              ||||||||||||||||||||||||||||||| |||||| |||||||||||||| ||||||
Db        443 ATCTCAACTGCACAACCCCTACTATGCCCCAGTTCAGCAGGAAGCAGTTAGAGTGGTCAT 384

Qy        244 CAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGG 303
              ||||||||||| |||||||||||||||||||||||||||||| |||||||||||||||||
Db        383 CAGCCAACCTCACCAACAGCACTTGGGTTTTCCTGTTGAGAGAGGGGACTGAGAGACAGG 324

Qy        304 ACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCA 363
              ||||||||||||||||||||| |  |||||||||||| |||||||||||||||||| |||
Db        323 ACTAGCTGGATTTCCTAGGCCGATTAAGAATCCCTAAACCTAGCTGGGAAGGTGACCGCA 264

Qy        364 TCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACT 423
              ||||||| |||||| ||||||||||||||||||||||||| |||||||||||||||||||
Db        263 TCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAACCAATCAGAGAGCTCACT 204

Qy        424 AAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGA 483
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        203 AAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGA 144

Qy        484 GCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAA 543
              |||||| ||||||||||||||| |  ||||||||||||||||||||||| || |  ||||
Db        143 GCACAGTGGGAGGGACAAGGATTGCAATATAAACCCAGGCATTCGAGCCAGC-ANAGCAA 85

Qy        544 CCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATT 603
              || |||||||||||| |||| ||||||||| |||||||||||||||||||||||||||||
Db         84 CCGCCTTTGGGTCCCTTCCCCTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCTATT 25

Qy        604 AAATCTTGCAACTGAAAAAAAAAA 627
              ||||||||||||||||||||||||
Db         24 AAATCTTGCAACTGAAAAAAAAAA 1


RESULT 28
AQ892947/c
LOCUS       AQ892947                 712 bp    DNA     linear   GSS 10-NOV-1999
DEFINITION  HS_3131_B2_E04_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3131 Col=8 Row=J, genomic survey
            sequence.
ACCESSION   AQ892947
VERSION     AQ892947.1  GI:6349137
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3131  row: J  column: 8
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 712.
```

```
FEATURES             Location/Qualifiers
     source          1..712
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3131 Col=8 Row=J"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

  Query Match             61.0%;  Score 387.2;  DB 8;  Length 712;
  Best Local Similarity   82.2%;  Pred. No. 7.2e-101;
  Matches  484;  Conservative    0;  Mismatches   94;  Indels   11;  Gaps    3;

Qy         56 AACTACAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTG 115
              || |   ||  |||  |  || ||| | |||     |||| || ||| | ||| |||  |
Db        699 AAGTTACAAAGGTTTCTNCAAAGGATCCCCAAGATCAGTCTAT-ACTCAAATCTACCCGG 641

Qy        116 GACCCCTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTC 175
                ||   ||||||||||||||| ||||||| ||||||| |||| ||  ||||||||||||
Db        640 ACCCTTGGGACCGGCCTGCTAGTCCATGCTTCGATGTTGATGATATCAAAGGCACCCCTC 581

Qy        176 CCGAGGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGA 235
              ||||||||||||||| ||||  |||| || | ||| ||| |||||| |||||||||||||
Db        580 CCGAGGAAATCTCAAGTGCATGACCCTTAGT-TGCACCAGTTCAGCAGGAAGCAGTTAGA 522

Qy        236 GCGGTCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGA 295
              |||| | |  |||||||||||||| || |||||||||||||||||||||||||   ||||
Db        521 GCGGCCGTTGGCCAACCTCCCCAATAGTACTTGGGTTTTCCTGTTGAGAGGGGTTGCTGA 462

Qy        296 GAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGG 355
              ||||||||||||||||||||||||||||| || |||||||||||||||||||||||||||
Db        461 GAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGGAAGG 402

Qy        356 TGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT----- 410
              ||||||||||||||| ||||||| ||||||||||| |||||||||||||||||||
Db        401 TGACTGCATCCACCTTTAAACACGGGGCTTGCAACGTAGCTCACACCCGACCAATGAGGT 342

Qy        411 ----CAGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAA 466
                   ||||||||||||||||||||||||||||||||| |||| ||||||||||||||||
Db        341 AGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAAGTAAAGAAATAGCCAA 282

Qy        467 TCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATT 526
              ||||||||  ||||||||||||| ||||||||||| |||| |||||||||||||||| ||
Db        281 TCATCTATCACCTGAGAGCACAGGGGGAGGGACAATGATCAGGATATAAACCCAGGCGTT 222

Qy        527 CGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCA 586
              | ||||||||||||| |||| |||||||| ||||||||||||||||| ||||||||||||
Db        221 CTAGCCGGCAACGGCTACCCTCTTTGGGTACCCTCCCTTTGTATGGGAGCTCTGTTTTCA 162

Qy        587 CTCTATTTCACTCTATTAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
              |||||||  |   |   ||  |    ||   |||  |||  |||  |||
Db        161 CTCTATTAAATCTTGCAAAGACACAAAAACCAAACCAAACCAAACCAAA 113


RESULT 29
AI128526/c
LOCUS       AI128526                 436 bp    mRNA    linear   EST 27-OCT-1998
DEFINITION  qc61h10.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA
            clone IMAGE:1714147 3' similar to contains PTR7.b1 PTR5 repetitive
            element ;, mRNA sequence.
ACCESSION   AI128526
VERSION     AI128526.1  GI:3597040
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 436)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Insert Length: 720   Std Error: 0.00
             Seq primer: -40m13 fwd. ET from Amersham
             High quality sequence stop: 428.
FEATURES              Location/Qualifiers
     source           1. .436
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1714147"
                      /dev_stage="two placentae: one from 8 weeks and another
                      from 9 weeks post conception"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
                      /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library constructed by Bento Soares and
                      M.Fatima Bonaldo."
ORIGIN

  Query Match             60.9%;  Score 386.8;  DB 1;  Length 436;
  Best Local Similarity   93.7%;  Pred. No. 8.4e-101;
  Matches  403;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy        206 TATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCAC 265
              || ||||||||||||| ||||||||||||||||||| || ||||||||||||||||||||
Db        436 TACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCAC 377

Qy        266 TTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCA 325
              || |||||||||||||||| |||||||||| |||||||||||||||||||||||||||
Db        376 TTAGGTTTTCCTGTTGAGATGGGGGACTGAAAGACAGGACTAGCTGGATTTCCTAGGCTG 317

Qy        326 ACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTT 385
              || |||||||||||||||||||||||||||||||  ||||||||| |||||| |||||||
Db        316 ACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTT 257

Qy        386 GCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAAT 445
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||| |
Db        256 GCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGAC 197

Qy        446 AGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGAT 505
              ||||||||||||||||||||||||||||||||||||||||||||| |||||||||| |||
Db        196 AGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGAT 137

Qy        506 CGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTT 565
              ||||||||||||||| |  |||||||||||||||||||||||||||||||||||||||||
Db        136 CGGGATATAAACCCAAGTTTTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTT 77

Qy        566 TGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTATTAAATCTTGCAACTGAAAAAAAA 625
              |||||||| ||||||||||||  |||||||||| |||||||||||||||||| |||||||
Db         76 TGTATGGGAGCTCTGTTTTCATGCTATTTCACTTTATTAAATCTTGCAACTGCAAAAAAA 17

Qy        626 AAAAAAAAAA 635
              ||||||||||
Db         16 AAAAAAAAAA 7


RESULT 30
AI393478/c
LOCUS       AI393478                 446 bp    mRNA    linear   EST 30-MAR-1999
DEFINITION  tg45g04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2111766 3', mRNA sequence.
ACCESSION   AI393478
```

```
VERSION     AI393478.1  GI:4223025
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1061   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 445.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2111766"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match             60.7%;  Score 385.2;  DB 1;  Length 446;
  Best Local Similarity   92.5%;  Pred. No. 2.5e-100;
  Matches  419;  Conservative    0;  Mismatches   24;  Indels   10;  Gaps    1;

Qy        183 AATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCA 242
              ||| |||||||  ||||||||| |||||||||||||||| ||||||||||| ||||||||
Db        446 AATTTCAACTGTGCAACCCCTANTATGCCCCAATTCAGCAGGAAGCAGTTACAGCGGTCA 387

Qy        243 TCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGACAG 302
              ||||||||||||||||||||||||||||||||||||||||||| ||| |||||||||||
Db        386 TCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGAGGGCACTGAGAGACAA 327

Qy        303 GACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACTGC 362
              |||||||||||||||||||||| || |||||||||||||| |||||||||||||||||||
Db        326 GACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCGTAGCTGGGAAGGTGACTGC 267

Qy        363 ATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCAC 422
              |||||||| |||||||||||||||| ||||||||||||||| ||||||||||||||||||
Db        266 ATCCACCTTTAAACATGGGGCTTGCCACTTAGCTCACACCCAACCAATCAGAGAGCTCAC 207

Qy        423 TAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAG 482
              |||||||||||||||||||||| |||||||||||||||||||||||| || |||||||||
Db        206 TAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCACCTTTTGCCTGAG 147

Qy        483 AGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGGCA 542
              |||||||| |||||||||||||||||||||||||||| ||||||| ||||||||| ||||
Db        146 AGCACAGCAGGAGGGACAAGGATCGGGATATAAACCCGGGCATTCAAGCCGGCAATGGCA 87

Qy        543 ACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCTAT 602
              ||||||||||||||||||||||||||||||| |||||||          |||||||||||
Db         86 ACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT----------TTTCACTCTAT 37

Qy        603 TAAATCTTGCAACTGAAAAAAAAAAAAAAAAAA 635
```

```
              ||||||||||||||| |||||||||||||||||
Db         36 TAAATCTTGCAACTGCAAAAAAAAAAAAAAAAA 4

Search completed: January 22, 2005, 19:31:32
Job time : 1902.54 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        January 22, 2005, 10:09:14 ; Search time 6212.46 Seconds
                                         (without alignments)
                                         11273.477 Million cell updates/sec

Title:         US-09-319-156B-9
Perfect score: 1481
Sequence:      1 atggccctcccttatcatac..........gtcttacaaatggaacccca 1481

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters:      9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
               1:  gb_ba:*
               2:  gb_htg:*
               3:  gb_in:*
               4:  gb_om:*
               5:  gb_ov:*
               6:  gb_pat:*
               7:  gb_ph:*
               8:  gb_pl:*
               9:  gb_pr:*
               10:  gb_ro:*
               11:  gb_sts:*
               12:  gb_sy:*
               13:  gb_un:*
               14:  gb_vi:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                   %
Result           Query
  No.   Score    Match Length DB  ID                       Description
-------------------------------------------------------------------------
    1    1481    100.0   1481  6  AR344387                 AR344387 Sequence
    2    1481    100.0   1481  6  AX001027                 AX001027 Sequence
    3    1481    100.0   1481  6  BD136197                 BD136197 Retrovira
    4    1481    100.0   1932 14   AF127228                 AF127228 Multiple
c   5  1468.2     99.1 162579  9  AL390039                 AL390039 Human DNA
    6  1466.6     99.0   2055  6  CQ719381                 CQ719381 Sequence
    7  1461.8     98.7   1629  6  BD252040                 BD252040 Method fo
    8  1461.8     98.7   1629  6  AX036992                 AX036992 Sequence
    9  1461.8     98.7   1629 14   AF331500                 AF331500 Multiple
   10  1461.8     98.7 172918  2  AC023366                 AC023366 Homo sapi
   11  1389.8     93.8   2074  6  AX478550                 AX478550 Sequence
   12  1341.8     90.6  10230  9  AY101589                 AY101589 Gorilla g
   13  1341.8     90.6 184675  9  AC145964                 AC145964 Pan trogl
   14  1340.2     90.5  10229  9  AY101586                 AY101586 Pan trogl
   15  1340.2     90.5  10229  9  AY101587                 AY101587 Pan trogl
   16  1340.2     90.5  10230  9  AY101588                 AY101588 Gorilla g
   17  1338.6     90.4  10246  9  AY101593                 AY101593 Hylobates
   18  1338.6     90.4  10248  9  AY101592                 AY101592 Hylobates
   19  1335.4     90.2   2694  9  AF520477S2               AF520478 Homo sapi
```

```
   20  1335.4     90.2   2694  9  AF520487S2               AF520488 Homo sapi
   21  1335.4     90.2   2694  9  AF520489S2               AF520490 Homo sapi
   22  1335.4     90.2   2694  9  AF520505S2               AF520506 Homo sapi
   23  1335.4     90.2   2694  9  AF520517S2               AF520518 Homo sapi
   24  1335.4     90.2   2694  9  AF520519S2               AF520520 Homo sapi
   25  1335.4     90.2   2694  9  AF520521S2               AF520522 Homo sapi
   26  1335.4     90.2   2694  9  AF520529S2               AF520530 Homo sapi
   27  1335.4     90.2   2694  9  AF520561S2               AF520562 Homo sapi
   28  1335.4     90.2   2930  6  AX355872                 AX355872 Sequence
   29  1335.4     90.2   2930  9  AF208161                 AF208161 Homo sapi
   30  1335.4     90.2   2946  6  AR177269                 AR177269 Sequence
   31  1335.4     90.2   2946  6  BD232464                 BD232464 Secreted
   32  1335.4     90.2   2946  6  BD086133                 BD086133 Secreted
   33  1333.8     90.1   1624  9  AF506835                 AF506835 Homo sapi
   34  1333.8     90.1   1860  9  AF513360                 AF513360 Homo sapi
   35  1333.8     90.1   2055  6  BD221827                 BD221827 Nucleic s
   36  1333.8     90.1   2055  6  AX007999                 AX007999 Sequence
   37  1333.8     90.1   2599  6  BD221806                 BD221806 Nucleic s
   38  1333.8     90.1   2599  6  AX007978                 AX007978 Sequence
   39  1333.8     90.1   2694  9  AF520481S2               AF520482 Homo sapi
   40  1333.8     90.1   2694  9  AF520483S2               AF520484 Homo sapi
   41  1333.8     90.1   2694  9  AF520485S2               AF520486 Homo sapi
   42  1333.8     90.1   2694  9  AF520491S2               AF520492 Homo sapi
   43  1333.8     90.1   2694  9  AF520493S2               AF520494 Homo sapi
   44  1333.8     90.1   2694  9  AF520495S2               AF520496 Homo sapi
   45  1333.8     90.1   2694  9  AF520497S2               AF520498 Homo sapi

                               ALIGNMENTS


RESULT 1
AR344387
LOCUS       AR344387                1481 bp    mRNA    linear   PAT 17-AUG-2003
DEFINITION  Sequence 105 from patent US 6582703.
ACCESSION   AR344387
VERSION     AR344387.1  GI:33740328
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1481)
  AUTHORS   Perron,H., Beseme,F., Bedin,F., Paranhos-Baccala,G.,
            Komurian-Pradel,F., Jolivet-Reynaud,C. and Mandrand,B.
  TITLE     Isolated nucleotide sequences associated with multiple sclerosis or
            rheumatoid arthritis and a process of detecting
  JOURNAL   Patent: US 6582703-A 105 24-JUN-2003;
FEATURES             Location/Qualifiers
     source          1..1481
                     /organism="unknown"
                     /mol_type="mRNA"
ORIGIN

  Query Match            100.0%;  Score 1481;  DB 6;  Length 1481;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
```

```
Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 2
AX001027
LOCUS       AX001027                1481 bp    DNA     linear   PAT 10-MAR-2000
DEFINITION  Sequence 9 from Patent WO9902666.
ACCESSION   AX001027
VERSION     AX001027.1  GI:7241264
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1481)
  AUTHORS   Ott,C. and Bedin,F.
  TITLE     RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR
            ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR
            DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
  JOURNAL   Patent: WO 9902666-A 9 21-JAN-1999;
            BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES             Location/Qualifiers
     source          1. .1481
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 1481;  DB 6;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 3
BD136197
LOCUS       BD136197                1481 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Retroviral nucleic material and nucleotide fragments, in
            particular, associated with multiple sclerosis and/or rheumatoid
            arthritis, for diagnostic, prophylactic and therapeutic uses.
ACCESSION   BD136197
VERSION     BD136197.1  GI:23231142
KEYWORDS    JP 2002509437-A/7.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1481)
  AUTHORS   Baccala,G.P., Pradel,F.K., Bedin,F., Sodoyer,M., Ott,C., Mallet,F.,
            Perron,H. and Mandrand,B.
  TITLE     Retroviral nucleic material and nucleotide fragments, in
            particular, associated with multiple sclerosis and/or rheumatoid
            arthritis, for diagnostic, prophylactic and therapeutic uses
  JOURNAL   Patent: JP 2002509437-A 7 26-MAR-2002;
            BIO MERIEUX
COMMENT     OS   Unidentified
            PN   JP 2002509437-A/7
            PD   26-MAR-2002
            PF   07-JUL-1998 JP 1999508255
            PR   07-JUL-1997 FR      97/08816
            PI   GLAUCIA PARAHNOS BACCALA,FLORENCE KOMURIAN PRADEL,FREDERIC  PI
              BEDIN,
            PI   MIREILLE SODOYER,CATHERINE OTT,FRANCOIS MALLET,HERVE PERRON,
            PI   BERNARD MANDRAND
            PC   C12N15/48,C12Q1/70,C07K14/15,A61K31/70
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   Retroviral nucleic material and nucleotide fragments, in  CC
                     particular,
            CC   associated with multiple sclerosis and/or
            rheumatoid arthritis,
            CC            for
            CC   diagnostic, prophylactic and therapeutic uses FH   Key
                   Location/Qualifiers
            FT   source             1. .1481
            FT                      /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1481
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 1481;  DB 6;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
```

```
RESULT 4
AF127228
LOCUS       AF127228                1932 bp    mRNA    linear   VRL 11-AUG-1999
DEFINITION  Multiple sclerosis associated retrovirus element clone CL15
          / env-like mRNA sequence.
ACCESSION   AF127228
VERSION     AF127228.1  GI:5726293
KEYWORDS    .
SOURCE      Multiple sclerosis associated retrovirus element
  ORGANISM  Multiple sclerosis associated retrovirus element
            Viruses; Retroid viruses; Retroviridae.
REFERENCE   1  (bases 1 to 1932)
  AUTHORS   Komurian-Pradel,F., Paranhos-Baccala,G., Bedin,F.,
            Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,
            Mallet,F., Mandrand,B. and Perron,H.
  TITLE     Molecular cloning and characterization of MSRV-related sequences
            associated with retrovirus-like particles
  JOURNAL   Virology 260 (1), 1-9 (1999)
  MEDLINE   99335590
   PUBMED   10405350
REFERENCE   2  (bases 1 to 1932)
  AUTHORS   Komurian-Pradel,F., Paranhos-Baccala,G., Bedin,F.,
            Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,
            Mallet,F., Mandrand,B. and Perron,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1999) UMR103 CNRS, bioMerieux, 46, allee
            D'Italie, Lyon 69007, France
FEATURES             Location/Qualifiers
     source          1. .1932
                     /organism="Multiple sclerosis associated retrovirus
                     element"
                     /mol_type="mRNA"
                     /db_xref="taxon:89382"
                     /clone="CL15"
     misc_feature    452. .1932
                     /note="similar to env but interrupted by premature stop
                     codon; includes signal peptide region, and partial SU and
                     TM regions"
     misc_difference 568
                     /note="in clones CL15-2 and CL37-14; results in a
                     premature stop codon to Trp amino acid translation"
                     /replace="g"
ORIGIN

  Query Match             100.0%;  Score 1481;  DB 14;  Length 1932;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        452 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 511

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 571

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 631

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        632 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 691

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        692 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 751

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        752 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 811
```

```
Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         812 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 871

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         872 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 931

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         932 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 991

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         992 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 1051

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1052 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1111

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1112 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1171

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1172 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 1231

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1232 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 1291

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1292 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1351

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1352 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 1411

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1412 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1471

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1472 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1531

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1532 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1591

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1592 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1651

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1652 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1711

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1712 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1771

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1772 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1831

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1832 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1891

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               |||||||||||||||||||||||||||||||||||||||||
Db        1892 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1932
```

```
RESULT 5
AL390039/c
LOCUS       AL390039              162579 bp    DNA     linear   PRI 22-DEC-2000
DEFINITION  Human DNA sequence from clone RP13-383K5 on chromosome Xq22.1-24
            Contains ESTs, STSs and GSSs. Contains a novel gene and a novel
            gene similar to NUP62 (nucleoporin 62kD), complete sequence.
ACCESSION   AL390039
VERSION     AL390039.10  GI:10186780
KEYWORDS    HTG; nucleoporin; NUP62.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162579)
  AUTHORS   Grafham,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2000 this sequence version replaced gi:9944073.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            IMPORTANT: This sequence is not the entire insert of clone
            RP13-383K5 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP13-383K5 is at 1 in this sequence. The
            true left end of clone RP5-1126E12 is at 162480 in this sequence.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            RP13-383K5 is from the library RPCI-13.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..162579
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q22.1-24"
                     /clone="RP13-383K5"
                     /clone_lib="RPCI-13.2"
     misc_feature    complement(1..232)
                     /note="match: GSS: Em:AQ036363"
     misc_feature    16..554
                     /note="match: GSS: Em:AQ504848"
     repeat_region   475..644
                     /note="AluSq repeat: matches 134..303 of consensus"
     repeat_region   742..919
                     /note="MLT1B repeat: matches 207..388 of consensus"
     repeat_region   1112..1142
                     /note="L1MEc repeat: matches 327..356 of consensus"
```

```
repeat_region   1143. .1456
                /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region   1457. .2938
                /note="L1MEc repeat: matches 356. .1897 of consensus"
repeat_region   2922. .3739
                /note="L1PA2 repeat: matches 5329. .6146 of consensus"
repeat_region   3844. .4087
                /note="L1M4 repeat: matches 1. .251 of consensus"
repeat_region   4092. .4231
                /note="L1M4 repeat: matches 612. .756 of consensus"
repeat_region   4373. .4559
                /note="L1M4c repeat: matches 971. .1152 of consensus"
repeat_region   4586. .4619
                /note="17 copies 2 mer aa 100% conserved"
repeat_region   4623. .4964
                /note="MLT1A1 repeat: matches 3. .333 of consensus"
repeat_region   4989. .5149
                /note="L1MA8 repeat: matches 6120. .6288 of consensus"
repeat_region   5152. .5602
                /note="L1PA15 repeat: matches 5325. .5779 of consensus"
repeat_region   5606. .5916
                /note="AluSq repeat: matches 1. .313 of consensus"
repeat_region   5917. .6014
                /note="49 copies 2 mer ga 75% conserved"
repeat_region   6018. .6409
                /note="L1PA15 repeat: matches 5769. .6157 of consensus"
repeat_region   6566. .6750
                /note="L1M4 repeat: matches 2857. .3036 of consensus"
repeat_region   6751. .7063
                /note="AluJo repeat: matches 1. .307 of consensus"
repeat_region   7064. .8097
                /note="L1M4 repeat: matches 3036. .4105 of consensus"
repeat_region   8099. .8419
                /note="AluSx repeat: matches 1. .306 of consensus"
repeat_region   8420. .8487
                /note="L1MC/D repeat: matches 5582. .5648 of consensus"
repeat_region   10039. .10396
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   10519. .10807
                /note="AluSq repeat: matches 3. .297 of consensus"
repeat_region   11343. .11416
                /note="37 copies 2 mer tt 67% conserved"
repeat_region   11448. .11983
                /note="L1MD3 repeat: matches 7162. .7739 of consensus"
repeat_region   12030. .12337
                /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region   12995. .13077
                /note="MIR repeat: matches 103. .186 of consensus"
repeat_region   13110. .13222
                /note="FLAM_C repeat: matches 2. .124 of consensus"
repeat_region   13228. .13518
                /note="L1MEc repeat: matches 2131. .2414 of consensus"
repeat_region   13519. .13853
                /note="MER7A repeat: matches 1. .346 of consensus"
repeat_region   13854. .14039
                /note="L1MEc repeat: matches 1951. .2131 of consensus"
repeat_region   14040. .14254
                /note="MER20 repeat: matches 1. .218 of consensus"
repeat_region   14255. .14807
                /note="L1MEc repeat: matches 1686. .1951 of consensus"
repeat_region   14808. .15105
                /note="AluJb repeat: matches 1. .311 of consensus"
repeat_region   15106. .15492
                /note="L1MEc repeat: matches 1309. .1686 of consensus"
repeat_region   15496. .15822
                /note="L1MB5 repeat: matches 5823. .6167 of consensus"
repeat_region   15824. .15920
                /note="L1MEc repeat: matches 1044. .1140 of consensus"
misc_feature    15885. .16086
                /note="match: STS: Em:G34444 Em:L14259"
misc_feature    15920. .16082
                /note="match: GSS: Em:AQ769824"
misc_feature    complement(15927. .16073)
                /note="match: GSS: Em:AQ054794 Em:AQ823559"
misc_feature    16067. .16294
                /note="match: GSS: Em:AQ457003"
repeat_region   16923. .16998
                /note="38 copies 2 mer at 71% conserved"
repeat_region   17107. .17166
                /note="L2 repeat: matches 2688. .2750 of consensus"
repeat_region   17175. .17755
                /note="L2 repeat: matches 1244. .1898 of consensus"
repeat_region   17839. .18106
                /note="MER70A repeat: matches 148. .418 of consensus"
repeat_region   18213. .18318
                /note="L2 repeat: matches 1128. .1247 of consensus"
repeat_region   18706. .18811
                /note="L2 repeat: matches 2631. .2750 of consensus"
repeat_region   18812. .19057
                /note="AluSx repeat: matches 1. .297 of consensus"
repeat_region   19058. .19492
                /note="L2 repeat: matches 2213. .2647 of consensus"
misc_feature    complement(19352. .19843)
                /note="match: GSS: Em:AQ340313"
repeat_region   19556. .19872
                /note="LTR17 repeat: matches 1. .326 of consensus"
repeat_region   19873. .21594
                /note="HERV17 repeat: matches 6812. .8523 of consensus"
repeat_region   21587. .22280
                /note="HERV17 repeat: matches 4301. .4980 of consensus"
repeat_region   22298. .22742
                /note="L1MB5 repeat: matches 5623. .6135 of consensus"
repeat_region   22743. .23048
                /note="AluY repeat: matches 1. .306 of consensus"
repeat_region   23049. .23286
                /note="L1MB5 repeat: matches 5403. .5623 of consensus"
repeat_region   23285. .23361
                /note="L1M4 repeat: matches 5006. .5082 of consensus"
repeat_region   23364. .23429
                /note="L1M4 repeat: matches 4019. .4084 of consensus"
repeat_region   23429. .24305
                /note="L1 repeat: matches 3097. .4043 of consensus"
repeat_region   24306. .24593
                /note="AluJo repeat: matches 1. .293 of consensus"
repeat_region   24594. .25109
                /note="L1 repeat: matches 2600. .3097 of consensus"
repeat_region   25110. .25436
                /note="L1M3 repeat: matches 5101. .5440 of consensus"
repeat_region   25437. .25742
                /note="AluSq repeat: matches 1. .307 of consensus"
repeat_region   25779. .25961
                /note="MIR repeat: matches 55. .260 of consensus"
repeat_region   25962. .26059
                /note="MIR repeat: matches 1. .89 of consensus"
repeat_region   26135. .26328
                /note="MLT1A1 repeat: matches 144. .352 of consensus"
repeat_region   26554. .26697
                /note="MLT1A1 repeat: matches 1. .149 of consensus"
repeat_region   26711. .26936
                /note="L1PA9 repeat: matches 5939. .6162 of consensus"
repeat_region   26981. .27154
                /note="L1MC/D repeat: matches 5620. .5805 of consensus"
repeat_region   27182. .27307
                /note="L1P4 repeat: matches 5454. .5584 of consensus"

Query Match             99.1%;  Score 1468.2;  DB 9;  Length 162579;
Best Local Similarity   99.5%;  Pred. No. 0;
Matches 1473;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      21559 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 21500

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21499 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 21440
```

```
Qy          121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21439 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 21380

Qy          181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21379 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 21320

Qy          241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21319 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 21260

Qy          301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21259 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 21200

Qy          361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21199 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 21140

Qy          421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21139 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 21080

Qy          481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21079 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 21020

Qy          541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21019 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 20960

Qy          601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20959 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 20900

Qy          661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20899 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 20840

Qy          721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20839 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 20780

Qy          781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20779 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 20720

Qy          841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20719 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 20660

Qy          901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20659 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 20600

Qy          961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20599 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 20540

Qy         1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20539 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 20480

Qy         1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20479 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 20420

Qy         1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20419 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 20360

Qy         1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20359 GGAGAAGAATGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 20300

Qy         1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        20299 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACACTGGGGCCTC 20240

Qy         1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20239 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 20180

Qy         1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20179 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 20120

Qy         1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
                ||||||||||||||||| |||||  | |||||||| |||||
Db        20119 GAAGCTGTAAAGCTACAAATGGTTCTTCAAATGGAGCCCCA 20079


RESULT 6
CQ719381
LOCUS       CQ719381                2055 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 5315 from Patent WO02068579.
ACCESSION   CQ719381
VERSION     CQ719381.1  GI:42280238
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 5315 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .2055
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             99.0%;  Score 1466.6;  DB 6;  Length 2055;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1472;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy            1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
                ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db           36 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 95

Qy           61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           96 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 155

Qy          121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          156 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 215

Qy          181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          216 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 275

Qy          241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          276 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 335

Qy          301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          336 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 395
```

```
Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 455

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 515

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        516 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 575

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 635

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        636 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 695

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        696 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 755

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        756 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 815

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        816 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 875

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        876 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 935

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        936 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 995

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        996 CTTCCTTTTGTTATCAAAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1055

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1056 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1115

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1116 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1175

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1176 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1235

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1236 GGAGAAGAATGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1295

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1296 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACACTGGGGCCTC 1355

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1356 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1415

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1416 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1475
```

```
Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||||||||| |||||  | |||||||| |||||
Db       1476 GAAGCTGTAAAGCTACAAATGGTTCTTCAAATGGAGCCCCA 1516

RESULT 7
BD252040
LOCUS       BD252040                1629 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Method for detecting superantigen activity in a biological sample.
ACCESSION   BD252040
VERSION     BD252040.1  GI:33061810
KEYWORDS    JP 2002539804-A/1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1629)
  AUTHORS   Perron,H. and Lafont,M.
  TITLE     Method for detecting superantigen activity in a biological sample
  JOURNAL   Patent: JP 2002539804-A 1 26-NOV-2002;
            BIO MERIEUX
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002539804-A/1
            PD   26-NOV-2002
            PF   20-MAR-2000 JP 2000607009
            PR   19-MAR-1999 FR     99/03622,28-OCT-1999 FR     99/13755 PI
            HERVE PERRON,MONIQUE LAFONT
            PC   C12N15/09,A61K31/711,A61K38/00,A61K39/395,A61K45/00,A61K48/00,
            PC   A61P25/00,
            PC   C07K14/47,C07K16/18,C12N7/00,C12Q1/02,C12Q1/68,G01N27/447, PC
             G01N33/15,
            PC
            G01N33/50,G01N33/569,G01N33/577//C12P21/08,(C12N7/00,C12R1:93), PC
             (C12Q1/02,C12R1:93),(C12Q1/68,C12R1:93),C12N15/00,G01N27/26, PC
            A61K37/02
            CC   Method for detecting superantigen activity in a biological  CC
                      sample
            FH   Key             Location/Qualifiers
            FT   source          1. .1629
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .1629
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             98.7%;  Score 1461.8;  DB 6;  Length 1629;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 1469;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy       481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy       541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy       601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy       661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy       721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy       781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy       841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy       901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy       961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy      1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy      1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy      1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy      1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy      1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy      1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
             |||||||||||||||||||||  |||||||||||||||||||||||||||| ||||||| 
Db      1321 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 1380

Qy      1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
             |||||||||||||||||||||||||| ||| |||||||||||||||||||||||||||||
Db      1381 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
```

```
Qy      1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
             ||||||||||||||||| || ||  | ||||||||||||||
Db      1441 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCA 1481


RESULT 8
AX036992
LOCUS       AX036992                1629 bp    DNA     linear   PAT 16-NOV-2000
DEFINITION  Sequence 1 from Patent FR2791140.
ACCESSION   AX036992
VERSION     AX036992.1  GI:11226420
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Lafont,M. and Perron,H.
  JOURNAL   Patent: FR 2791140-A 1 22-SEP-2000;
            BIO MERIEUX (FR)
FEATURES             Location/Qualifiers
     source          1..1629
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             98.7%;  Score 1461.8;  DB 6;  Length 1629;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 1469;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy         1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy        61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 120

Qy       121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy       181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy       241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy       301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy       481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy       541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy       601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy          661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy          721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy          781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy          841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy          901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy          961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy         1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy         1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy         1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy         1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy         1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy         1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
                |||||||||||||||||||||  |||||||||||||||||||||||||||| ||||||| 
Db         1321 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 1380

Qy         1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||| ||| |||||||||||||||||||||||||||||
Db         1381 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy         1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
                ||||||||||||||||| || ||  | ||||||||||||||
Db         1441 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCA 1481
```

```
RESULT 9
AF331500
LOCUS       AF331500                1629 bp    mRNA    linear   VRL 21-NOV-2001
DEFINITION  Multiple sclerosis associated retrovirus element clone pV14
            recombinant envelope protein (env) mRNA, complete cds.
ACCESSION   AF331500
VERSION     AF331500.1  GI:13310190
KEYWORDS    .
SOURCE      Multiple sclerosis associated retrovirus element
  ORGANISM  Multiple sclerosis associated retrovirus element
            Viruses; Retroid viruses; Retroviridae.
REFERENCE   1  (bases 1 to 1629)
  AUTHORS   Perron,H., Jouvin-Marche,E., Michel,M., Ounanian-Paraz,A.,
            Camelo,S., Dumon,A., Jolivet-Reynaud,C., Marcel,F., Souillet,Y.,
            Borel,E., Gebuhrer,L., Santoro,L., Marcel,S., Seigneurin,J.M.,
            Marche,P.N. and Lafon,M.
  TITLE     Multiple sclerosis retrovirus particles and recombinant envelope
            trigger an abnormal immune response in vitro, by inducing
            polyclonal Vbeta16 T-lymphocyte activation
  JOURNAL   Virology 287 (2), 321-332 (2001)
  MEDLINE   21425392
   PUBMED   11531410
REFERENCE   2  (bases 1 to 1629)
  AUTHORS   Perron,H., Ounanian-Paraz,A., Marcel,F. and Souillet,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2000) R&D, bioMerieux, Chemin de L'Orme, Marcy
            L'Etoile 69280, France
FEATURES             Location/Qualifiers
     source          1. .1629
                     /organism="Multiple sclerosis associated retrovirus
                     element"
                     /mol_type="mRNA"
                     /db_xref="taxon:89382"
                     /clone="pV14"
     gene            1. .1629
                     /gene="env"
     CDS             1. .1629
                     /gene="env"
                     /function="displaying superantigen-like properties causing
                     polyclonal activation of human Vbeta 16 T-cells (CD3+)
                     independent from the T-cell receptor antigen specificity"
                     /note="surface and transmembrane regions"
                     /codon_start=1
                     /product="recombinant envelope protein"
                     /protein_id="AAK18189.1"
                     /db_xref="GI:13310191"
                     /translation="MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLWRTRLP
                     GNIDAPSYRSLSKGNSTFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGA
                     TVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRT
                     HTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSA
                     YHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLG
                     TGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAK
                     RGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWMPWTLP
                     FLGPLAAIIFLLLFGPCIFNFLVKFVSSRIEAVKLQIVLQMEPQMQSMTKIYRGPLDR
                     PARLCSDVNDIEVTPPEEISTAQPLLHSNSVGSS"
ORIGIN

  Query Match             98.7%;  Score 1461.8;  DB 14;  Length 1629;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 1469;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy            1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy           61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db           61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 120

Qy          121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy          181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy          241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy          301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
```

```
Qy       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy       481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy       541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy       601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy       661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy       721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy       781 ATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy       841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy       901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy       961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy      1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy      1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy      1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy      1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy      1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy      1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
             |||||||||||||||||||||  |||||||||||||||||||||||||||| |||||||
Db      1321 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 1380

Qy      1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
             |||||||||||||||||||||||||| ||| |||||||||||||||||||||||||||||
Db      1381 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy      1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
```

```
             ||||||||||||||||| || ||  | | ||||||||||||||
Db      1441 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCA 1481

RESULT 10
AC023366
LOCUS       AC023366               172918 bp    DNA     linear   HTG 02-MAR-2000
DEFINITION  Homo sapiens clone RP11-15N10, WORKING DRAFT SEQUENCE, 33 unordered
            pieces.
ACCESSION   AC023366
VERSION     AC023366.2  GI:7143443
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172918)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-15N10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172918)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 2, 2000 this sequence version replaced gi:6970363.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3492
                Center clone name: 15_N_10
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 142569 bases at least Q40
                Consensus quality: 156863 bases at least Q30
                Consensus quality: 164107 bases at least Q20
                Insert size: 178000; agarose-fp
                Insert size: 169718; sum-of-contigs
                Quality coverage: 2.9 in Q20 bases; agarose-fp
                Quality coverage: 3.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
```

```
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *        1      1266: contig of 1266 bp in length
           *     1267      1366: gap of 100 bp
           *     1367      2686: contig of 1320 bp in length
           *     2687      2786: gap of 100 bp
           *     2787      4288: contig of 1502 bp in length
           *     4289      4388: gap of 100 bp
           *     4389      5572: contig of 1184 bp in length
           *     5573      5672: gap of 100 bp
           *     5673      6812: contig of 1140 bp in length
           *     6813      6912: gap of 100 bp
           *     6913      7404: contig of 492 bp in length
           *     7405      7504: gap of 100 bp
           *     7505      8915: contig of 1411 bp in length
           *     8916      9015: gap of 100 bp
           *     9016     10200: contig of 1185 bp in length
           *    10201     10300: gap of 100 bp
           *    10301     12003: contig of 1703 bp in length
           *    12004     12103: gap of 100 bp
           *    12104     14038: contig of 1935 bp in length
           *    14039     14138: gap of 100 bp
           *    14139     16226: contig of 2088 bp in length
           *    16227     16326: gap of 100 bp
           *    16327     19115: contig of 2789 bp in length
           *    19116     19215: gap of 100 bp
           *    19216     21492: contig of 2277 bp in length
           *    21493     21592: gap of 100 bp
           *    21593     24794: contig of 3202 bp in length
           *    24795     24894: gap of 100 bp
           *    24895     27624: contig of 2730 bp in length
           *    27625     27724: gap of 100 bp
           *    27725     30780: contig of 3056 bp in length
           *    30781     30880: gap of 100 bp
           *    30881     34253: contig of 3373 bp in length
           *    34254     34353: gap of 100 bp
           *    34354     38260: contig of 3907 bp in length
           *    38261     38360: gap of 100 bp
           *    38361     42757: contig of 4397 bp in length
           *    42758     42857: gap of 100 bp
           *    42858     46232: contig of 3375 bp in length
           *    46233     46332: gap of 100 bp
           *    46333     51049: contig of 4717 bp in length
           *    51050     51149: gap of 100 bp
           *    51150     53373: contig of 2224 bp in length
           *    53374     53473: gap of 100 bp
           *    53474     58170: contig of 4697 bp in length
           *    58171     58270: gap of 100 bp
           *    58271     63398: contig of 5128 bp in length
           *    63399     63498: gap of 100 bp
           *    63499     69283: contig of 5785 bp in length
           *    69284     69383: gap of 100 bp
           *    69384     75177: contig of 5794 bp in length
           *    75178     75277: gap of 100 bp
           *    75278     80817: contig of 5540 bp in length
           *    80818     80917: gap of 100 bp
           *    80918     89553: contig of 8636 bp in length
           *    89554     89653: gap of 100 bp
           *    89654    101099: contig of 11446 bp in length
           *   101100    101199: gap of 100 bp
           *   101200    113135: contig of 11936 bp in length
           *   113136    113235: gap of 100 bp
           *   113236    132520: contig of 19285 bp in length
           *   132521    132620: gap of 100 bp
           *   132621    150646: contig of 18026 bp in length
           *   150647    150746: gap of 100 bp
           *   150747    172918: contig of 22172 bp in length.
FEATURES             Location/Qualifiers
     source          1..172918
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-15N10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1266
                     /note="assembly_fragment"
     misc_feature    1367..2686
                     /note="assembly_fragment"
     misc_feature    2787..4288
                     /note="assembly_fragment"
     misc_feature    4389..5572
                     /note="assembly_fragment"
     misc_feature    5673..6812
                     /note="assembly_fragment"
     misc_feature    6913..7404
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    7505..8915
                     /note="assembly_fragment"
     misc_feature    9016..10200
                     /note="assembly_fragment"
     misc_feature    10301..12003
                     /note="assembly_fragment"
     misc_feature    12104..14038
                     /note="assembly_fragment"
     misc_feature    14139..16226
                     /note="assembly_fragment"
     misc_feature    16327..19115
                     /note="assembly_fragment"
     misc_feature    19216..21492
                     /note="assembly_fragment"
     misc_feature    21593..24794
                     /note="assembly_fragment"
     misc_feature    24895..27624
                     /note="assembly_fragment"
     misc_feature    27725..30780
                     /note="assembly_fragment"
     misc_feature    30881..34253
                     /note="assembly_fragment"
     misc_feature    34354..38260
                     /note="assembly_fragment"
     misc_feature    38361..42757
                     /note="assembly_fragment"
     misc_feature    42858..46232
                     /note="assembly_fragment"
     misc_feature    46333..51049
                     /note="assembly_fragment"
     misc_feature    51150..53373
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    53474..58170
                     /note="assembly_fragment"
     misc_feature    58271..63398
                     /note="assembly_fragment"
     misc_feature    63499..69283
                     /note="assembly_fragment"
     misc_feature    69384..75177
                     /note="assembly_fragment"
     misc_feature    75278..80817

  Query Match             98.7%;  Score 1461.8;  DB 2;  Length 172918;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 1469;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||||| |||||||||||||||||||
Db     103234 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTCACCCCCTTTCGCTCTCACT 103293

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     103294 GCACCGCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATAAAGA 103353
```

```
Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103354 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 103413

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103414 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 103473

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103474 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 103533

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103534 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 103593

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103594 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 103653

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103654 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 103713

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103714 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 103773

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103774 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 103833

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103834 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 103893

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103894 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 103953

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103954 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 104013

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104014 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 104073

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104074 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 104133

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104134 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 104193

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db     104194 CTTCCTTTTGTTATCAAAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 104253

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104254 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 104313

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104314 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 104373

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104374 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 104433

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
```

```
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104434 GGAGAAGAATGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 104493

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     104494 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACACTGGGGCCTC 104553

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104554 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 104613

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104614 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 104673

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||||||||| |||||  | |||||||| |||||
Db     104674 GAAGCTGTAAAGCTACAAATGGTTCTTCAAATGGAGCCCCA 104714


RESULT 11
AX478550
LOCUS       AX478550                2074 bp    DNA     linear   PAT 12-AUG-2002
DEFINITION  Sequence 4 from Patent WO0248362.
ACCESSION   AX478550
VERSION     AX478550.1  GI:22217319
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ramkumar,J. and Arvizu,C.
  TITLE     Embryogenesis associated proteins
  JOURNAL   Patent: WO 0248362-A 4 20-JUN-2002;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2074
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 7477736CB1"
ORIGIN

  Query Match             93.8%;  Score 1389.8;  DB 6;  Length 2074;
  Best Local Similarity   96.2%;  Pred. No. 0;
  Matches 1424;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              |||||||||||||||  || |||||||||||||||||||| ||| ||||||  |||||||
Db         64 ATGGCCCTCCCTTATTGTATTTTTCTCTTTACTGTTCTCTCACCACCTTTCAGTCTCACT 123

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              |||||| ||||||||  |||||  |||||||||||||||||||||| ||||||||| |||
Db        124 GCACCCTCTCCATGCCACTGTAGGACCAGTAGCTCCCCTTACCAAGCGTTTCTATGGAGA 183

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | ||||| |||  || |||||||||||||||| |||||||||||||||||||||||| ||
Db        184 ATGCGGCGTCCCAGACATATTGATGCCCCATCGTATAGGAGTTTATCTAAGGGAAACCCC 243

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||| ||||||||||||||||||| ||||||||||||||||||||||||||| |||
Db        244 GCCTTCACCGCCCACACCCATATGCCCCACAACTGCTATAACTCTGCCACTCTTTGTATG 303

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 363

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||| |||||||||||||||||||||||||||| |||||||||||||||||||| ||
Db        364 GGAGCCACTATCTGTTGGACTTACTTCACCCATACCGGTATGTCTGATGGGGGTGGAGTT 423

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
```

```
              |||| |||||||||||||||||| |||||||||| |||||||||||| ||||||| ||||
Db        424 CAAGATCAGGCAAGAGAAAAACATGTAAAGGAAGTAATCTCCCAACTCACCCGGGTACAT 483

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        484 AGCACCCCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 543

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||| |||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        544 CATACTCACCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 603

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        604 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAGTC 663

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| |||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db        664 CCTGTACCTGAACAATGGAATAACTTCAGCACAGAAATAAACACCACTTCCATTTTAGTA 723

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        724 GGACCTCTTGTTTCCAATCTGGAAATAATCCATACCTCAAACCTCACCTGTGTAAAATTT 783

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||||||  ||||||| |||||| |||||||| |||||||| |||||||||| |
Db        784 AGCAATACTATAGCAACAACCAACTCCCAGTGCATCAGATGGGTAACTCCTCCCACACAA 843

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 903

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 963

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 1023

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1083

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1143

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1203

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1263

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1264 GGAGAAGAATGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1323

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1324 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACACTGGGGCCTC 1383

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1384 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1443

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1444 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1503

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||||||||| |||||  | |||||||| |||||
Db       1504 GAAGCTGTAAAGCTACAAATGGTTCTTCAAATGGAGCCCCA 1544
```

```
RESULT 12
AY101589
LOCUS       AY101589               10230 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Gorilla gorilla isolate 1 endogenous retrovirus ERV-W, ERVWE1
            locus, allele B, complete sequence.
ACCESSION   AY101589
VERSION     AY101589.1  GI:37544413
KEYWORDS    .
SOURCE      Gorilla gorilla (gorilla)
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 10230)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 10230)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .10230
                     /organism="Gorilla gorilla"
                     /mol_type="genomic DNA"
                     /isolate="1"
                     /isolation_source="PBMC of gorilla 1"
                     /db_xref="taxon:9593"
                     /cell_type="PBMC"
                     /note="isolated by PCR; ERVWE1 locus, allele B
                     endogenous_virus: ERV-W"
     LTR             1. .795
     mRNA            join(248. .826,2894. .9821)
                     /note="putative mRNA transcript 1; gag-pol pseudogene"
     mRNA            join(248. .826,7595. .9821)
                     /product="envelope glycoprotein"
                     /note="putative mRNA transcript 2"
     mRNA            join(248. .826,9258. .9821)
                     /note="putative mRNA transcript 3; no coding region"
     CDS             7822. .9438
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAM68168.1"
                     /db_xref="GI:37544414"
                     /translation="MALPYHILLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMRRP
                     GNIDAPSHRSFSKGTPTFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTHTGMSDGGGVQDQAREKHVKEVISQLTRVHSTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLDFRPYVSIPVPEEWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTIDTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVVSKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKCTPPEEISTAQPLLRPNSAGSS"
     LTR             9497. .10230
ORIGIN

  Query Match             90.6%;  Score 1341.8;  DB 9;  Length 10230;
  Best Local Similarity   94.1%;  Pred. No. 0;
  Matches 1394;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| | |||||||||||||||| |||||| ||||| ||||||||
Db       7822 ATGGCCCTCCCTTATCATATTCTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 7881
```

```
Qy       61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
            ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db     7882 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 7941

Qy      121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
            | ||||| ||||||||||||||||||||||||  |||||||||| |||||||||| | ||
Db     7942 ATGCGGCGTCCTGGAAATATTGATGCCCCATCGCATAGGAGTTTTTCTAAGGGAACCCCC 8001

Qy      181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8002 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 8061

Qy      241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
            |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     8062 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 8121

Qy      301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
            |||| ||||||||||||||||||||||||||||||  |||||||||||||||||||| ||
Db     8122 GGAGTCACTGTCTGTTGGACTTACTTCACCCATACTGGTATGTCTGATGGGGGTGGAGTT 8181

Qy      361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
            |||| |||||||||||||||||| ||||| |||| |||||||||||||||||||| ||||
Db     8182 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTGACCCGGGTACAT 8241

Qy      421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
            |||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     8242 AGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 8301

Qy      481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
            |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db     8302 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 8361

Qy      541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
            ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db     8362 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGGACTTCAGGCCATATGTTTCAATC 8421

Qy      601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
            ||||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     8422 CCTGTACCTGAAGAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 8481

Qy      661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8482 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 8541

Qy      721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
            |||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||| |
Db     8542 AGCAATACTATAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 8601

Qy      781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
            ||||||||||||||||||||||||||||| |||||||||||||||||||||| |||||||
Db     8602 ATAGTCTGCCTACCCTCAGGAATATTTTTCGTCTGTGGTACCTCAGCCTATCGTTGTTTG 8661

Qy      841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
            |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db     8662 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCTCCTATGACCATCTAC 8721

Qy      901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
            |||||||||||||||||||| ||||||||| ||||||||| |||||||||||||||||||
Db     8722 ACTGAACAAGATTTATACAATTATGTCGTATCTAAGCCCCGCAACAAAAGAGTACCCATT 8781

Qy      961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
            ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db     8782 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 8841

Qy     1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
            ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db     8842 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 8901

Qy     1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
            ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8902 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 8961

Qy     1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
            ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db     8962 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 9021

Qy     1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
            || |||||| |||||||||||||||||||||||||||| |||| ||||||||||||||||
Db     9022 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 9081

Qy     1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
            |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db     9082 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 9141

Qy     1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
            ||||||||||||||||||||| |||||||||||||||||||||||| ||||||| |||||||
Db     9142 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTTGCAGCTATAATATTG 9201

Qy     1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
             |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db     9202 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 9261

Qy     1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
            ||||||||||| ||||| ||||      | ||| |  || |
Db     9262 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 9302


RESULT 13
AC145964
LOCUS       AC145964               184675 bp    DNA     linear   PRI 29-OCT-2003
DEFINITION  Pan troglodytes BAC clone RP43-12F2 from 7, complete sequence.
ACCESSION   AC145964
VERSION     AC145964.3  GI:36016769
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 184675)
  AUTHORS   Shahid,S., Cotton,M., Bielicki,L. and Meyer,R.
  TITLE     The sequence of Pan troglodytes BAC clone RP43-12F2
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 184675)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Sequencing of Pan troglodytes
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 184675)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 184675)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 184675)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 184675)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 26, 2003 this sequence version replaced gi:35073486.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: C_PT012F02
            --------------.
            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
            was isolated from white blood cells obtained from a male chimpanzee
            (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
            clone and detailed information can be obtained from ResGen
            (http://www.resgen.com) or Pieter de Jong and co-workers at
            http://www.bacpac.chori.org.

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..184675
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /map="7"
                     /clone="RP43-12F2"
                     /clone_lib="RPCI-43"
     unsure          115419..115500
                     /note="Sequence derived from one plasmid subclone."
     unsure          138694..138701
                     /note="Sequence derived from one plasmid subclone."
     misc_feature    140277..140388
                     /note="Sequence derived from PCR product of project DNA."
ORIGIN

  Query Match             90.6%;  Score 1341.8;  DB 9;  Length 184675;
  Best Local Similarity   94.1%;  Pred. No. 0;
  Matches 1394;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db     153623 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 153682

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db     153683 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 153742

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||||||||||||||| |||||||||| | ||
Db     153743 ATGCAGCGTCCCGGAAATATTGATGCCCCATCATATAGGAGTTTTTCTAAGGGAACCCCC 153802

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     153803 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 153862

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     153863 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 153922

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||||||  |||||||||||||||||||| ||
Db     153923 GGAGTCACTGTCTGTTGGACTTACTTCACCCATACTGGTATGTCTGATGGGGGTGGAGTT 153982

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db     153983 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 154042

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              |||||| |||| |||||||||||||||| |||||||||||||||||||| ||||||||||
Db     154043 AGCACCTCTAGGCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAAACCTCCGTACC 154102

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db     154103 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 154162

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db     154163 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 154222

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154223 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 154282

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154283 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 154342

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| ||||||||||| |||||||||||||||||||||||| |||||||||| |
Db     154343 AGCAATACTACAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 154402

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     154403 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 154462

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154463 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 154522

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              |||||||||||||||||||| ||||||||| ||||||||| |||||||||||||||||||
Db     154523 ACTGAACAAGATTTATACAATTATGTCGTATCTAAGCCCCGCAACAAAAGAGTACCCATT 154582

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db     154583 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 154642

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db     154643 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 154702

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     154703 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 154762

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db     154763 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 154822

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db     154823 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 154882

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db     154883 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 154942

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db     154943 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 155002

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db     155003 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 155062
```

```
Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db     155063 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 155103


RESULT 14
AY101586
LOCUS       AY101586               10229 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Pan troglodytes isolate 1 endogenous retrovirus ERV-W, ERVWE1
            locus, allele A, complete sequence.
ACCESSION   AY101586
VERSION     AY101586.1  GI:37544407
KEYWORDS    .
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 10229)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 10229)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .10229
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /isolate="1"
                     /isolation_source="PBMC of chimpanzee 1"
                     /db_xref="taxon:9598"
                     /cell_type="PBMC"
                     /note="isolated by PCR; ERVWE1 locus, allele A
                     endogenous_virus: ERV-W"
     LTR             1. .779
     mRNA            join(248. .810,2889. .9820)
                     /note="putative mRNA transcript 1; gag-pol pseudogene"
     mRNA            join(248. .810,7592. .9820)
                     /product="envelope glycoprotein"
                     /note="putative mRNA transcript 2"
     mRNA            join(248. .810,9255. .9820)
                     /note="putative mRNA transcript 3; no coding region"
     CDS             7819. .9435
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAM68165.1"
                     /db_xref="GI:37544408"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSFSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKIINPSCPGGLGV
                     TVCWTYFTHTGMSDGGGVQDQAREKHVKEVISQLTRVHSASRPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTDTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVVSKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEILTAQPLLRPNSAGSS"
     LTR             9494. .10229
ORIGIN

  Query Match            90.5%;  Score 1340.2;  DB 9;  Length 10229;
  Best Local Similarity  94.1%;  Pred. No. 0;
  Matches 1393;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
```

```
Db       7819 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 7878

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       7879 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 7938

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||||||||||||||| |||||||||| | ||
Db       7939 ATGCAGCGTCCCGGAAATATTGATGCCCCATCATATAGGAGTTTTTCTAAGGGAACCCCC 7998

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       7999 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 8058

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| ||||| |||||||||||||||||||||||||||
Db       8059 CATGCAAATACTCATTATTGGACAGGAAAAATAATTAATCCTAGTTGTCCTGGAGGACTT 8118

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||||||  |||||||||||||||||||| ||
Db       8119 GGAGTCACTGTCTGTTGGACTTACTTCACCCATACTGGTATGTCTGATGGGGGTGGAGTT 8178

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       8179 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 8238

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||| || |||| |||||||||||||||| |||||||||||||||||||||||||||||||
Db       8239 AGCGCCTCTAGGCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 8298

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       8299 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 8358

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       8359 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 8418

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8419 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 8478

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8479 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 8538

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| ||||||||||| |||||||||||||||||||||||| |||||||||| |
Db       8539 AGCAATACTACAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 8598

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       8599 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 8658

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8659 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 8718

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| |||||||| ||||||||| |||||||||||||||||||
Db       8719 ACTGAACAAGATTTATACAATTATGTCGTATCTAAGCCCCGCAACAAAAGAGTACCCATT 8778

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       8779 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 8838

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       8839 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 8898

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8899 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 8958
```

```
Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       8959 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 9018

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       9019 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 9078

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       9079 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 9138

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       9139 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 9198

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db       9199 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 9258

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       9259 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 9299


RESULT 15
AY101587
LOCUS       AY101587               10229 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Pan troglodytes isolate 1 endogenous retrovirus ERV-W, ERVWE1
            locus, allele B, complete sequence.
ACCESSION   AY101587
VERSION     AY101587.1  GI:37544409
KEYWORDS    .
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 10229)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 10229)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1..10229
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /isolate="1"
                     /isolation_source="PBMC of chimpanzee 1"
                     /db_xref="taxon:9598"
                     /cell_type="PBMC"
                     /note="isolated by PCR; ERVWE1 locus, allele B
                     endogenous_virus: ERV-W"
     LTR             1..779
     mRNA            join(248..810,2889..9820)
                     /note="putative mRNA transcript 1; gag-pol pseudogene"
     mRNA            join(248..810,7592..9820)
                     /product="envelope glycoprotein"
                     /note="putative mRNA transcript 2"
     mRNA            join(248..810,9255..9820)
                     /note="putative mRNA transcript 3; no coding region"
     CDS             7819..9435
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAM68166.1"
                     /db_xref="GI:37544410"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSFSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKIINPSCPGGLGV
                     TVCWTYFTHTGMSDGGGVQDQAREKHVKEVISQLTRVHSASRPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTDTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVVSKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEILTAQPLLRPNSAGSS"
     LTR             9494..10229
ORIGIN

  Query Match            90.5%;  Score 1340.2;  DB 9;  Length 10229;
  Best Local Similarity  94.1%;  Pred. No. 0;
  Matches 1393;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db       7819 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 7878

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       7879 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 7938

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||||||||||||||| |||||||||| | ||
Db       7939 ATGCAGCGTCCCGGAAATATTGATGCCCCATCATATAGGAGTTTTTCTAAGGGAACCCCC 7998

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       7999 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 8058

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| ||||| |||||||||||||||||||||||||||
Db       8059 CATGCAAATACTCATTATTGGACAGGAAAAATAATTAATCCTAGTTGTCCTGGAGGACTT 8118

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||||||  |||||||||||||||||||| ||
Db       8119 GGAGTCACTGTCTGTTGGACTTACTTCACCCATACTGGTATGTCTGATGGGGGTGGAGTT 8178

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       8179 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 8238

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||| || |||| |||||||||||||||| |||||||||||||||||||||||||||||||
Db       8239 AGCGCCTCTAGGCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 8298

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       8299 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 8358

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       8359 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 8418

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8419 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 8478

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8479 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 8538

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| |||||||||||| ||||||||||||||||||||||| |||||||||| |
Db       8539 AGCAATACTACAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 8598

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        8599 ATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 8658

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8659 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 8718

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||||| |||||||| ||||||||| |||||||||||||||||||
Db        8719 ACTGAACAAGATTTATACAATTATGTCGTATCTAAGCCCCGCAACAAAAGAGTACCCATT 8778

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        8779 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 8838

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        8839 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 8898

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8899 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 8958

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        8959 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 9018

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db        9019 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 9078

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db        9079 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 9138

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        9139 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 9198

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db        9199 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 9258

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||| ||||| ||||      | ||| |  || |
Db        9259 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 9299


RESULT 16
AY101588
LOCUS       AY101588                10230 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Gorilla gorilla isolate 1 endogenous retrovirus ERV-W, ERVWE1
            locus, allele A, complete sequence.
ACCESSION   AY101588
VERSION     AY101588.1  GI:37544411
KEYWORDS    .
SOURCE      Gorilla gorilla (gorilla)
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 10230)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 10230)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
```

```
     source          1. .10230
                     /organism="Gorilla gorilla"
                     /mol_type="genomic DNA"
                     /isolate="1"
                     /isolation_source="PBMC of gorilla 1"
                     /db_xref="taxon:9593"
                     /cell_type="PBMC"
                     /note="isolated by PCR; ERVWE1 locus, allele A
                     endogenous_virus: ERV-W"
     LTR             1. .795
     mRNA            join(248. .826,2894. .9821)
                     /note="putative mRNA transcript 1; gag-pol pseudogene"
     mRNA            join(248. .826,7595. .9821)
                     /product="envelope glycoprotein"
                     /note="putative mRNA transcript 2"
     mRNA            join(248. .826,9258. .9821)
                     /note="putative mRNA transcript 3; no coding region"
     CDS             7822. .9438
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAM68167.1"
                     /db_xref="GI:37544412"
                     /translation="MALPYHILLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMRRP
                     GNIDAPSHRSFSKGTPTFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTHTGMSDGGGVQDQAREKHVKEVISQLTRVHSTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLDFRPYVSIPVPEEWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTIDTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVVSKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKCTPPEEISTAQPLLRPNSAGSS"
     LTR             9497. .10230
ORIGIN

  Query Match             90.5%;  Score 1340.2;  DB 9;  Length 10230;
  Best Local Similarity   94.1%;  Pred. No. 0;
  Matches 1393;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| | |||||||||||||||| |||||| ||||| ||||||||
Db        7822 ATGGCCCTCCCTTATCATATTCTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 7881

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        7882 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 7941

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | ||||| ||||||||||||||||||||||||  |||||||||| |||||||||| | ||
Db        7942 ATGCGGCGTCCTGGAAATATTGATGCCCCATCGCATAGGAGTTTTTCTAAGGGAACCCCC 8001

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8002 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 8061

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        8062 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 8121

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||||||  |||||||||||||||||||| ||
Db        8122 GGAGTCACTGTCTGTTGGACTTACTTCACCCATACTGGTATGTCTGATGGGGGTGGAGTT 8181

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||||||||||| ||||
Db        8182 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTGACCCGGGTACAT 8241

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               |||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        8242 AGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 8301

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
```

```
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       8302 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 8361

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       8362 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGGACTTCAGGCCATATGTTTCAATC 8421

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       8422 CCTGTACCTGAAGAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 8481

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8482 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 8541

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||||||||||||||| |||||||||||||||||||||||| |||||||||| |
Db       8542 AGCAATACTATAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 8601

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||| |||||||||||||||||||||| |||||||
Db       8602 ATAGTCTGCCTACCCTCAGGAATATTTTTCGTCTGTGGTACCTCAGCCTATCGTTGTTTG 8661

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db       8662 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCTCCTATGACCATCTAC 8721

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| |||||||| ||||||||| |||||||||||||||||||
Db       8722 ACTGAACAAGATTTATACAATTATGTCGTATCTAAGCCCCGCAACAAAAGAGTACCCATT 8781

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  |||| |||||||||||   ||||||||||||||||||| |||||
Db       8782 CTTCCTTTTGTTATAGGAGCGGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 8841

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       8842 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 8901

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8902 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 8961

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       8962 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 9021

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       9022 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 9081

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       9082 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 9141

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| |||||||||||||||||||||| |||||| ||||||||
Db       9142 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTTGCAGCTATAATATTG 9201

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       9202 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 9261

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       9262 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 9302


RESULT 17
AY101593
LOCUS       AY101593               10246 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Hylobates pileatus isolate 1 endogenous retrovirus ERV-W, ERVWE1
            locus, allele B, complete sequence.
ACCESSION   AY101593
VERSION     AY101593.1  GI:37544421
KEYWORDS    .
SOURCE      Hylobates pileatus (pileated gibbon)
  ORGANISM  Hylobates pileatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
REFERENCE   1  (bases 1 to 10246)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 10246)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .10246
                     /organism="Hylobates pileatus"
                     /mol_type="genomic DNA"
                     /isolate="1"
                     /isolation_source="PBMC of gibbon 1"
                     /db_xref="taxon:9589"
                     /cell_type="PBMC"
                     /note="isolated by PCR; ERVWE1 locus, allele B
                     endogenous_virus: ERV-W"
     LTR             1. .773
     mRNA            join(248. .254,2843. .9807)
                     /note="putative mRNA transcript 1; gag-pol pseudogene"
     mRNA            join(248. .254,7579. .9807)
                     /product="envelope glycoprotein"
                     /note="putative mRNA transcript 2"
     mRNA            join(248. .254,9242. .9807)
                     /note="putative mRNA transcript 3; no coding region"
     CDS             7806. .9422
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAM68172.1"
                     /db_xref="GI:37544422"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRTQRP
                     GNIDAPLYRSFSKGSPTFTAHTYMPRTCYNSATLCMHANTQYWTGKMINPSCPGGLGV
                     TVCWTYFTHTGMSDGGGVQDQAREKHVKEVISQLTQVHSTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWMCLPLDFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTDTTNSQCIRWVTPPTRIFCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVVSKPRNKRVPILPFVMGAGVLGALG
                     TGIGSITTSTQFYYKLSRELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNIGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAIKLQMEPKMESKTKNYRRSLDWPASP
                     RSDVNDIKGIPPEEISTAQPLLRPNSAGSS"
     LTR             9481. .10246
ORIGIN

  Query Match             90.4%;  Score 1338.6;  DB 9;  Length 10246;
  Best Local Similarity   94.0%;  Pred. No. 0;
  Matches 1392;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| ||||||||||||||||||||||||| ||||| ||||||||
Db       7806 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCTCTTTCACTCTCACT 7865

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       7866 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 7925

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              |||| || |||||||||||||||||||||||  ||||||||||| |||||||||| | ||
Db       7926 ACGCAGCGTCCTGGAAATATTGATGCCCCATTGTATAGGAGTTTTTCTAAGGGAAGCCCC 7985
```

```
Qy       181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
             |||||||||||||||||| |||||||||||| ||||||||||||||||||||||||||||
Db      7986 ACCTTCACTGCCCACACCTATATGCCCCGCACCTGCTATAACTCTGCCACTCTTTGCATG 8045

Qy       241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
             |||||||||||||| ||||||||||| |||||||||||||||||||||||||||||||||
Db      8046 CATGCAAATACTCAGTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 8105

Qy       301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
             |||| ||||||||||||||||||||||||||||||  |||||||||||||||||||| ||
Db      8106 GGAGTCACTGTCTGTTGGACTTACTTCACCCATACTGGTATGTCTGATGGGGGTGGAGTT 8165

Qy       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
             |||| |||||||||||||||||| ||||| |||| ||||||||||||||||| || ||||
Db      8166 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTGACCCAGGTACAT 8225

Qy       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
             |||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      8226 AGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 8285

Qy       481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
             || ||||||||||| ||||||||||||||||| |||||| |||||||||||||||| |||
Db      8286 CACACTCGCCTGGTAAGCCTATTTAATACCACACTCACTGGGCTCCATGAGGTCTCGGCC 8345

Qy       541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
             |||||||||||||||||||||||||||||||||||| |||||||||||||  ||||||||
Db      8346 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGGACTTCAGGCCATATGTTTCAATC 8405

Qy       601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
             ||||| ||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      8406 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAATACCACTTCCGTTTTAGTA 8465

Qy       661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
             ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      8466 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACTTCAAACCTCACCTGTGTAAAATTT 8525

Qy       721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
             |||||||||| ||||||||||| |||||||||||||||||||||||| ||||||||||||
Db      8526 AGCAATACTACAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACGA 8585

Qy       781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
             ||| |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      8586 ATATTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 8645

Qy       841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8646 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 8705

Qy       901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
             ||||||||||||||||||||| |||||||| ||||||||| |||||||||||||||||||
Db      8706 ACTGAACAAGATTTATACAATTATGTCGTATCTAAGCCCCGCAACAAAAGAGTACCCATT 8765

Qy       961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
             ||||||||||||||  ||||||||||||||||   |||||||||||||||||||||||||
Db      8766 CTTCCTTTTGTTATGGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCAGTATC 8825

Qy      1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
             |||||||||||||||||||||||||||||||||| |||| |||||||||||||||||| |
Db      8826 ACAACCTCTACTCAGTTCTACTACAAACTATCTCGAGAACTAAATGGTGACATGGAACGG 8885

Qy      1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
             ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8886 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 8945

Qy      1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
             ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db      8946 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 9005

Qy      1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
             || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db      9006 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 9065

Qy      1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
             |||||||||||||||||| |||||||||||||||||| |||||  | || ||||||||||
Db      9066 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACATTGGACCCTGGGGCCTC 9125

Qy      1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
             ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db      9126 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 9185

Qy      1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              | |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      9186 CTGCTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATT 9245

Qy      1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
             |||||| |||| ||||| ||||      | |||||  || |
Db      9246 GAAGCTATAAAACTACAAATGGAGCCCAAGATGGAGTCCAA 9286
```

```
RESULT 18
AY101592
LOCUS       AY101592               10248 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Hylobates pileatus isolate 1 endogenous retrovirus ERV-W, ERVWE1
            locus, allele A, complete sequence.
ACCESSION   AY101592
VERSION     AY101592.1  GI:37544419
KEYWORDS    .
SOURCE      Hylobates pileatus (pileated gibbon)
  ORGANISM  Hylobates pileatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
REFERENCE   1  (bases 1 to 10248)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 10248)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .10248
                     /organism="Hylobates pileatus"
                     /mol_type="genomic DNA"
                     /isolate="1"
                     /isolation_source="PBMC of gibbon 1"
                     /db_xref="taxon:9589"
                     /cell_type="PBMC"
                     /note="isolated by PCR; ERVWE1 locus, allele A
                     endogenous_virus: ERV-W"
     LTR             1. .773
     mRNA            join(248. .804,2997. .9809)
                     /note="putative mRNA transcript 1; gag-pol pseudogene"
     mRNA            join(248. .804,7581. .9809)
                     /product="envelope glycoprotein"
                     /note="putative mRNA transcript 2"
     mRNA            join(248. .804,9244. .9809)
                     /note="putative mRNA transcript 3; no coding region"
     CDS             7808. .9424
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAM68171.1"
                     /db_xref="GI:37544420"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRTQRP
                     GNIDAPLYRSFSKGSPTFTAHTYMPRTCYNSATLCMHANTQYWTGKMINPSCPGGLGV
                     TVCWTYFTHTGMSDGGGVQDQAREKHVKEVISQLTQVHSTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWMCLPLDFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTDTTNSQCIRWVTPPTRIFCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVVSKPRNKRVPILPFVMGAGVLGALG
                     TGIGSITTSTQFYYKLSRELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNIGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAIKLQMEPKMESKTKNYRRSLDWPASP
                     RSDVNDIKGIPPEEISTAQPLLRPNSAGSS"
     LTR             9483. .10248
ORIGIN

  Query Match             90.4%;  Score 1338.6;  DB 9;  Length 10248;
  Best Local Similarity   94.0%;  Pred. No. 0;
  Matches 1392;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| ||||||||||||||||||||||||| ||||| ||||||||
Db       7808 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCTCTTTCACTCTCACT 7867

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       7868 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 7927

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              |||| || |||||||||||||||||||||||  ||||||||||| |||||||||| | ||
Db       7928 ACGCAGCGTCCTGGAAATATTGATGCCCCATTGTATAGGAGTTTTTCTAAGGGAAGCCCC 7987

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              |||||||||||||||||| |||||||||||| ||||||||||||||||||||||||||||
Db       7988 ACCTTCACTGCCCACACCTATATGCCCCGCACCTGCTATAACTCTGCCACTCTTTGCATG 8047

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||| ||||||||||| |||||||||||||||||||||||||||||||||
Db       8048 CATGCAAATACTCAGTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 8107

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||||||  |||||||||||||||||||| ||
Db       8108 GGAGTCACTGTCTGTTGGACTTACTTCACCCATACTGGTATGTCTGATGGGGGTGGAGTT 8167

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| ||||||||||||||||| || ||||
Db       8168 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTGACCCAGGTACAT 8227

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              |||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       8228 AGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 8287

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              || ||||||||||| ||||||||||||||||| |||||| |||||||||||||||| |||
Db       8288 CACACTCGCCTGGTAAGCCTATTTAATACCACACTCACTGGGCTCCATGAGGTCTCGGCC 8347

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              |||||||||||||||||||||||||||||||||||| |||||||||||||  ||||||||
Db       8348 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGGACTTCAGGCCATATGTTTCAATC 8407

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       8408 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAATACCACTTCCGTTTTAGTA 8467

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       8468 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACTTCAAACCTCACCTGTGTAAAATTT 8527

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| ||||||||||| |||||||||||||||||||||||| ||||||||||||
Db       8528 AGCAATACTACAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACGA 8587

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||| |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       8588 ATATTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 8647

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8648 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 8707

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| |||||||| ||||||||| |||||||||||||||||||
Db       8708 ACTGAACAAGATTTATACAATTATGTCGTATCTAAGCCCCGCAACAAAAGAGTACCCATT 8767

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   |||||||||||||||||||||||||
Db       8768 CTTCCTTTTGTTATGGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCAGTATC 8827

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              |||||||||||||||||||||||||||||||||| |||| |||||||||||||||||| |
Db       8828 ACAACCTCTACTCAGTTCTACTACAAACTATCTCGAGAACTAAATGGTGACATGGAACGG 8887

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8888 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 8947

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       8948 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 9007

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       9008 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 9067

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||  | || ||||||||||
Db       9068 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACATTGGACCCTGGGGCCTC 9127

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       9128 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 9187

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               | |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       9188 CTGCTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATT 9247

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||| |||| ||||| ||||      | |||||  || |
Db       9248 GAAGCTATAAAACTACAAATGGAGCCCAAGATGGAGTCCAA 9288


RESULT 19
AF520477S2
LOCUS       AF520477S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 132 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520478
VERSION     AF520478.1  GI:33410924
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 132 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17561.1"
                     /db_xref="GI:33410925"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2694;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        284 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 343

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        344 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 403

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        404 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 463

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db        464 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 523

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        524 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 583

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        584 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 643

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        644 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 703

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        704 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 763

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        764 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 823

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        824 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 883

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        884 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 943
```

```
Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        944 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1003

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       1004 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1063

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1064 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1123

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1124 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1183

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1184 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1243

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1244 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1303

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1304 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1363

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1364 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1423

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1424 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1483

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       1484 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       1544 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1603

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       1604 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1663

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1664 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1723

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       1724 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1764
```

```
RESULT 20
AF520487S2
LOCUS       AF520487S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 148 allele B, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520488
VERSION     AF520488.1  GI:33410944
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 148 allele B"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17566.1"
                     /db_xref="GI:33410945"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2694;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        284 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 343

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        344 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 403

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        404 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 463

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        464 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 523

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        524 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 583

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        584 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 643
```

```
Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        644 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 703

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        704 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 763

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        764 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 823

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        824 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 883

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        884 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 943

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        944 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1003

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       1004 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1063

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1064 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1123

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1124 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1183

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1184 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1243

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1244 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1303

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1304 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1363

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1364 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1423

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1424 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1483

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       1484 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       1544 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1603

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       1604 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1663

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db       1664 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1723

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
```

```
                ||||||||||| ||||| ||||      | ||| |  || |
Db          1724 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1764

RESULT 21
AF520489S2
LOCUS       AF520489S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 22 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520490
VERSION     AF520490.1  GI:33410948
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 22 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17567.1"
                     /db_xref="GI:33410949"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2694;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy             1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
                 ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db           284 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 343

Qy            61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
                 ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db           344 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 403

Qy           121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
                 | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db           404 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 463

Qy           181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
                 ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db           464 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 523

Qy           241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
                 |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db           524 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 583

Qy           301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
                 |||| |||||||||||||||||||||||||||| ||  ||||||||||||||||||| ||
Db           584 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 643

Qy           361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
                 |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db           644 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 703

Qy           421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                  ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db           704 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 763

Qy           481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
                 |||||||||||||| ||||||||||||||||||||||||| |||||||||||||||| |||
Db           764 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 823

Qy           541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
                 ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db           824 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 883

Qy           601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 660
                 ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           884 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 943

Qy           661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           944 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1003

Qy           721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
                 |||||||||| | ||||||||| ||||||||||||||||||||||| |||||||||| |
Db          1004 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1063

Qy           781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
                 |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db          1064 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1123

Qy           841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1124 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1183

Qy           901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
                 ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db          1184 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1243

Qy           961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
                 ||||||||||||||  ||||||||||||||||   |||||||||||||||||||| |||||
Db          1244 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1303

Qy          1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
                 ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db          1304 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1363

Qy          1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
                 ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1364 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1423

Qy          1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
```

```
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1424 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1483

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       1484 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       1544 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1603

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       1604 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1663

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db       1664 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1723

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       1724 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1764


RESULT 22
AF520505S2
LOCUS       AF520505S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 37 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520506
VERSION     AF520506.1  GI:33410980
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 37 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17575.1"
                     /db_xref="GI:33410981"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2694;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        284 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 343

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        344 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 403

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        404 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 463

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        464 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 523

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        524 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 583

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        584 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 643

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        644 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 703

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        704 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 763

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        764 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 823

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        824 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 883

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        884 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 943

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        944 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1003

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       1004 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1063

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1064 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1123

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1124 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1183

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1184 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1243

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1244 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1303

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1304 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1363

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1364 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1423

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1424 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1483

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       1484 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       1544 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1603

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       1604 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1663

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1664 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1723

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       1724 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1764


RESULT 23
AF520517S2
LOCUS       AF520517S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 55 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520518
VERSION     AF520518.1  GI:33411004
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 55 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17581.1"
                     /db_xref="GI:33411005"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2694;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        284 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 343

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        344 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 403

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        404 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 463

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        464 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 523

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        524 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 583

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        584 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 643

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        644 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 703

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        704 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 763

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        764 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 823

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        824 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 883
```

```
Qy          601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
                ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          884 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 943

Qy          661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          944 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1003

Qy          721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
                |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db         1004 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1063

Qy          781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
                |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db         1064 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1123

Qy          841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1124 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1183

Qy          901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
                ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db         1184 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1243

Qy          961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
                ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db         1244 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1303

Qy         1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
                ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db         1304 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1363

Qy         1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
                ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1364 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1423

Qy         1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
                ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db         1424 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1483

Qy         1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
                || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db         1484 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543

Qy         1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
                |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db         1544 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1603

Qy         1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
                ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db         1604 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1663

Qy         1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                 |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db         1664 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1723

Qy         1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
                ||||||||||| ||||| ||||      | ||| |  || |
Db         1724 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1764


RESULT 24
AF520519S2
LOCUS       AF520519S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 55 allele B, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520520
VERSION     AF520520.1  GI:33411008
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 55 allele B"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17582.1"
                     /db_xref="GI:33411009"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2694;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy            1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
                ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db          284 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 343

Qy           61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
                ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db          344 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 403

Qy          121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
                | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db          404 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 463

Qy          181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
                ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db          464 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 523

Qy          241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
                |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db          524 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 583
```

```
Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| |||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        584 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 643

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| ||||||||||| ||||||| ||||
Db        644 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 703

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| |||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        704 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 763

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| ||||||||||||||||||||||||| |||||||||||||||| |||
Db        764 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 823

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| ||||||||||||||  ||||||||
Db        824 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 883

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        884 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 943

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        944 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1003

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| ||||||||||||||||||||||||| |||||||||| |
Db       1004 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1063

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1064 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1123

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1124 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1183

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              |||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1184 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1243

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   |||||||||||||||||||| |||||
Db       1244 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1303

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1304 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1363

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1364 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1423

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1424 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1483

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| ||||||||||||||||||||||
Db       1484 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| ||||||||||||||||| ||||| | ||  ||||||||||
Db       1544 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1603

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db       1604 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1663

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1664 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1723

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       1724 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1764

RESULT 25
AF520521S2
LOCUS       AF520521S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 6 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520522
VERSION     AF520522.1  GI:33411012
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 6 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17583.1"
                     /db_xref="GI:33411013"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2694;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;
```

```
Qy      1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
          ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db    284 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 343

Qy     61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
          ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db    344 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 403

Qy    121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
          | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db    404 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 463

Qy    181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    464 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 523

Qy    241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    524 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 583

Qy    301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
          |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db    584 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 643

Qy    361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
          |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db    644 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 703

Qy    421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
           ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    704 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 763

Qy    481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
          |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db    764 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 823

Qy    541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
          ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db    824 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 883

Qy    601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    884 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 943

Qy    661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    944 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1003

Qy    721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
          |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db   1004 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1063

Qy    781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   1064 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1123

Qy    841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1124 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1183

Qy    901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
          ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db   1184 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1243

Qy    961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
          ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db   1244 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1303

Qy   1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
          ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db   1304 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1363

Qy   1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
          ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1364 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1423

Qy   1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
          ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db   1424 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1483

Qy   1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
          || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db   1484 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543

Qy   1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
          |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db   1544 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1603

Qy   1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
          ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db   1604 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1663

Qy   1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   1664 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1723

Qy   1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
          ||||||||||| ||||| ||||      | ||| |  || |
Db   1724 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1764
```

```
RESULT 26
AF520529S2
LOCUS       AF520529S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 78 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520530
VERSION     AF520530.1  GI:33411028
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 78 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17587.1"
                     /db_xref="GI:33411029"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705..1720
                     /note="splice acceptor site"
     LTR             1959..2694
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2694;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        284 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 343

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        344 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 403

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        404 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 463

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        464 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 523

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        524 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 583

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        584 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 643

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        644 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 703

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        704 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 763

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        764 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 823

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        824 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 883

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        884 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 943

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        944 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1003

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       1004 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1063

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1064 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1123

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1124 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1183

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1184 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1243

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1244 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1303

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1304 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1363

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1364 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1423

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1424 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1483

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       1484 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       1544 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1603

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       1604 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1663

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1664 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1723

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       1724 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1764


RESULT 27
AF520561S2
LOCUS       AF520561S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 71 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520562
VERSION     AF520562.1  GI:33411092
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
```

```
           d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 71 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="African
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17603.1"
                     /db_xref="GI:33411093"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2694;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db         284 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 343

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         344 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 403

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db         404 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 463

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         464 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 523

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db         524 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 583

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db         584 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 643

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db         644 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 703

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         704 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 763

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db         764 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 823

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| ||||||||||||| ||||||||||||  ||||||||
Db         824 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 883

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         884 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 943

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         944 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1003

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db        1004 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1063

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        1064 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1123

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1124 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1183

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db        1184 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1243

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        1244 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1303

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1304 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1363

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1364 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1423

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        1424 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1483

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db        1484 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db        1544 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1603

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        1604 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1663

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db        1664 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1723

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||| ||||| ||||      | ||| |  || |
Db        1724 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1764


RESULT 28
AX355872
LOCUS       AX355872                2930 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 1 from Patent WO0204678.
ACCESSION   AX355872
```

```
VERSION     AX355872.1  GI:18620523
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Keith,J.C., Mccoy,J.M. and Mi,S.
  TITLE     Methods and compositions for diagnosing and treating preeclampsia
            and gestational trophoblast disorders
  JOURNAL   Patent: WO 0204678-A 1 17-JAN-2002;
            GENETICS INSTITUTE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .2930
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             930. .2546
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD23000:1"
                     /db_xref="GI:18620524"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
ORIGIN

  Query Match            90.2%;  Score 1335.4;  DB 6;  Length 2930;
  Best Local Similarity  93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        930 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 989

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        990 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1049

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db       1050 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1109

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1110 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1169

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1170 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1229

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1230 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1289

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       1290 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1349

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1350 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1409

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1410 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1469
```

```
Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1470 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1529

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 1589

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1649

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | |||||||||| ||||||||||||||||||||||| |||||||||| |
Db       1650 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1709

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1710 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1769

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1770 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1829

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1830 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1889

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1890 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1949

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1950 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2009

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2010 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2069

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       2070 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2129

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       2130 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2189

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       2190 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2249

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       2250 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2309

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db       2310 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2369

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       2370 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2410


RESULT 29
AF208161
LOCUS       AF208161              2930 bp    mRNA    linear   PRI 22-FEB-2000
DEFINITION  Homo sapiens syncytin precursor, mRNA, complete cds.
ACCESSION   AF208161
VERSION     AF208161.1  GI:6760400
KEYWORDS    .
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2930)
  AUTHORS   Mi,S., Lee,X., Li,X., Veldman,G.M., Finnerty,H., Racie,L.,
            LaVallie,E., Tang,X.Y., Edouard,P., Howes,S., Keith,J.C. Jr. and
            McCoy,J.M.
  TITLE     Syncytin is a captive retroviral envelope protein involved in human
            placental morphogenesis
  JOURNAL   Nature 403 (6771), 785-789 (2000)
  MEDLINE   20155476
   PUBMED   10693809
REFERENCE   2  (bases 1 to 2930)
  AUTHORS   Sha,M., Lee,X., Li,X., Veldman,G.M., Finnerty,H., Racie,L.,
            LaVallie,E., Tang,X., Edouard,P., Howes,S., Keith,J.C. Jr. and
            McCoy,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1999) Genetics Institute, 87 Cambridge Park
            Drive, Cambridge, MA 02140, USA
FEATURES             Location/Qualifiers
     source          1. .2930
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="AJ172-2s"
                     /tissue_type="testis"
     repeat_region   1. .2930
                     /rpt_family="Human endogeneous retrovirus HERV-W"
     5'UTR           1. .929
     CDS             930. .2546
                     /note="envelope protein"
                     /codon_start=1
                     /product="syncytin precursor"
                     /protein_id="AAF28334.1"
                     /db_xref="GI:6760401"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     sig_peptide     930. .989
     mat_peptide     990. .2543
                     /product="syncytin"
     misc_feature    1869. .1880
                     /note="furin; cleavage site"
                     /evidence=not_experimental
     misc_feature    2043. .2120
                     /note="Region: immunosuppresive region"
     misc_feature    2235. .2357
                     /note="transmembrane-region site"
     3'UTR           2547. .2930
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2930;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        930 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 989

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        990 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1049

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db       1050 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1109

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1110 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1169

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1170 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1229

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1230 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1289

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       1290 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1349

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1350 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1409

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1410 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1469

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1470 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1529

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1589

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1649

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       1650 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1709

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1710 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1769

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1770 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1829

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1830 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1889

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1890 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1949

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1950 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2009

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2010 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2069

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       2070 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2129

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       2130 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2189
```

```
Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||||||||||||||||| ||||||||||||||||||| |||||||| | || ||||||||||
Db        2190 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2249

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               |||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        2250 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2309

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||| 
Db        2310 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2369

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAA ATGGAACCCCA 1481
               ||||||||||| ||||| ||||      | ||| |  || |
Db        2370 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2410


RESULT 30
AR177269
LOCUS       AR177269                2946 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 3 from patent US 6312921.
ACCESSION   AR177269
VERSION     AR177269.1  GI:17919624
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2946)
  AUTHORS   Jacobs,K., McCoy,J.M., LaVallie,E.R., Racie,L.A., Evans,C.,
            Merberg,D., Mi,S. and Treacy,M.
  TITLE     Secreted proteins and polynucleotides encoding them
  JOURNAL   Patent: US 6312921-A 3 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1..2946
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             90.2%;  Score 1335.4;  DB 6;  Length 2946;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| |||||| ||||||||
Db         928 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 987

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         988 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1047

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        1048 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1107

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        1108 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1167

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        1168 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1227

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  ||||||||||||||||||||| ||
Db        1228 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1287

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| ||||||||||||||||||| ||||| |||| |||||||||||||| ||||||| ||||
Db        1288 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1347

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| |||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        1348 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1407

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| ||||||||||||||||||||||||| |||||||||||||||| |||
Db        1408 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1467

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||||| |||||||||||||  ||||||||
Db        1468 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1527

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1528 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1587

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1588 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1647

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||||||||| | ||||||||||| |||||||||||||||||||||||| |||||||||| |
Db        1648 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1707

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        1708 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1767

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1768 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1827

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||||| ||||| || ||||||||| ||||||||||||||||||||
Db        1828 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1887

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   |||||||||||||||||||| |||||
Db        1888 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1947

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1948 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2007

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2008 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2067

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        2068 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2127

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| |||||||||||||||||||||||||| |||| ||||||||||||||||||||
Db        2128 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2187

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||||||||||||||||| ||||||||||||||||||| |||||||| | || ||||||||||
Db        2188 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2247

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               |||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        2248 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2307

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||| 
Db        2308 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2367

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||| ||||| ||||      | ||| |  || |
Db        2368 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2408


Search completed: January 22, 2005, 16:13:53
Job time : 6221.46 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        January 22, 2005, 11:11:14 ; Search time 679.669 Seconds
                                           (without alignments)
                                           11438.492 Million cell updates/sec

Title:         US-09-319-156B-9
Perfect score: 1481
Sequence:      1 atggccctcccttatcatac..........gtcttacaaatggaacccca 1481

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters:      8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_23Sep04:*
               1:  geneseqn1980s:*
               2:  geneseqn1990s:*
               3:  geneseqn2000s:*
               4:  geneseqn2001as:*
               5:  geneseqn2001bs:*
               6:  geneseqn2002as:*
               7:  geneseqn2002bs:*
               8:  geneseqn2003as:*
               9:  geneseqn2003bs:*
               10:  geneseqn2003cs:*
               11:  geneseqn2003ds:*
               12:  geneseqn2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                 %
Result           Query
  No.    Score   Match Length DB  ID                          Description
  ---------------------------------------------------------------------------
    1     1481   100.0   1481  2  AAV43217                    Aav43217 Multiple
    2     1481   100.0   1481  2  AAX29703                    Aax29703 Clone C15
    3     1481   100.0   1481 12  ADG14846                     Adg14846 MSRV asso
    4     1481   100.0   2030  3  AAA63826                    Aaa63826 Nucleotid
    5   1461.8    98.7   1629  3  AAA96625                    Aaa96625 DNA encod
    6     1450    97.9   1481  9  ADB84400                    Adb84400 MSRV-1 as
    7   1389.8    93.8   2074  6  AAD41225                    Aad41225 Human EMB
    8   1335.4    90.2   2930  6  AAD24195                    Aad24195 Human syn
    9   1335.4    90.2   2946  2  AAX77526                    Aax77526 Human sec
   10   1335.4    90.2   2946  3  AAZ59468                    Aaz59468 Human sec
   11   1335.4    90.2   2946 10  ADC38776                     Adc38776 Human cDN
   12   1333.8    90.1   1617  5  AAH20070                    Aah20070 HERV-W en
   13   1333.8    90.1   2055  3  ABN97948                    Abn97948 Human ret
   14   1333.8    90.1   2599  3  ABN97927                    Abn97927 Human ret
   15   1333.8    90.1  10499  3  ABN97929                    Abn97929 Human ret
   16   1333.8    90.1  56093  6  ABL61744                    Abl61744 Colon ade
   17   1330.6    89.8   2781  5  AAF55630                    Aaf55630 Nucleotid
   18   1324.2    89.4   7582  3  AAA59215                    Aaa59215 Human end
   19   1322.6    89.3   7582  2  AAX25665                    Aax25665 Complete
   20   1308.2    88.3   2782  5  AAH20069                    Aah20069 HERV-W en
   21   1306.6    88.2   2782  2  AAX25661                    Aax25661 Human end
```

```
   22  1306.6    88.2   2782  3  AAA59211                Aaa59211 5' non co
   23  1298.2    87.7   3464 10  ADE09587                 Ade09587 Novel DNA
   24  1298.2    87.7   9502 10  ADF59718                 Adf59718 Human con
   25    1282    86.6  46340  3  ABN97978                Abn97978 Human ret
c  26  1278.8    86.3   1799  3  ABN97931                Abn97931 Human ret
   27    1239    83.7   2784  3  ABN97930                Abn97930 Human ret
   28  1131.6    76.4   1894  4  AAI14608                Aai14608 Probe #45
   29  1131.6    76.4   1894  4  ABA56337                Aba56337 Human foe
   30  1131.6    76.4   1894  4  AAI35980                Aai35980 Probe #46
   31  1131.6    76.4   1894  4  ABA45822                Aba45822 Human bre
   32  1131.6    76.4   1894  4  ABA25978                Aba25978 Probe #44
   33  1131.6    76.4   1894  4  AAK30018                Aak30018 Human bon
   34  1131.6    76.4   1894  4  AAK04516                Aak04516 Human bra
   35  1131.6    76.4   1894  4  ABS29670                Abs29670 Human liv
   36  1131.6    76.4   1894  5  AAI04422                Aai04422 Probe #44
   37  1131.6    76.4   1894  6  ABS04589                Abs04589 Human gen
   38  1116.4    75.4   1948  2  AAX25659                Aax25659 Human end
   39  1116.4    75.4   1948  3  AAA59209                Aaa59209 5' non co
   40    1048    70.8   6394  5  AAS84210                Aas84210 DNA encod
   41   716.6    48.4    792  4  AAI23803                Aai23803 Probe #13
   42   716.6    48.4    792  4  ABA68919                Aba68919 Human foe
   43   716.6    48.4    792  4  AAI49115                Aai49115 Probe #17
   44   716.6    48.4    792  4  ABA50940                Aba50940 Human bre
   45   716.6    48.4    792  4  ABA35872                Aba35872 Probe #14

                              ALIGNMENTS

RESULT 1
AAV43217
ID   AAV43217 standard; cDNA; 1481 BP.
XX
AC   AAV43217;
XX
DT   29-DEC-1998  (first entry)
XX
DE   Multiple sclerosis associated retrovirus fragment 5.
XX
KW   Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW   env gene; rheumatoid arthritis-associated virus; ss.
XX
OS   Multiple sclerosis associated retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1479
FT                   /*tag=  a
FT                   /product= "Encodes protein AAW71068"
FT                   /transl_except= (pos:115-117, appears to code for a stop
FT                   codon)
FT                   /note= "CDS does not contain a stop codon"
XX
PN   WO9823755-A1.
XX
PD   04-JUN-1998.
XX
PF   26-NOV-1997;   97WO-IB001482.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B;
XX
DR   WPI; 1998-322732/28.
DR   P-PSDB; AAW71068.
XX
PT   New nucleic acid from retroviruses - useful for diagnosis, prevention and
PT   treatment of, e.g. multiple sclerosis.
XX
PS   Disclosure; Page 184-185; 286pp; English.
XX
```

```
CC   The present sequence represents a multiple sclerosis (MS) associated
CC   retrovirus (MSRV) genomic fragment used in the method of the invention.
CC   The invention provides complete or partial genomic sequences of the MSRV-
CC   1 pol gene, gag gene and env gene, and polypeptides encoded by these
CC   genes. The invention also provides antibodies raised against the
CC   polypeptides. The genomic sequences, polypeptides and antibodies are also
CC   claimed useful for diagnosing infection by MS and rheumatoid arthritis-
CC   associated viruses, and also for prevention and treatment of infection
CC   with these viruses
XX
SQ   Sequence 1481 BP; 412 A; 410 C; 261 G; 398 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1481;  DB 2;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 2
AAX29703
ID   AAX29703 standard; DNA; 1481 BP.
XX
AC   AAX29703;
XX
DT   17-OCT-2003  (revised)
DT   27-AUG-2003  (revised)
DT   08-JUN-1999  (first entry)
XX
DE   Clone C15 from MSRV-1.
XX
KW   Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KW   rheumatoid polyarthritis; ss.
XX
OS   Viruses.
XX
PN   FR2765588-A1.
XX
PD   08-JAN-1999.
XX
PF   07-JUL-1997;   97FR-00008816.
XX
PR   07-JUL-1997;   97FR-00008816.
XX
PA   (INMR ) BIO MERIEUX.
XX
DR   WPI; 1999-098275/09.
DR   P-PSDB; AAW99553.
```

```
XX
PT   Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT   multiple sclerosis or rheumatoid polyarthritis.
XX
PS   Claim 1; Page 37-38; 83pp; French.
XX
CC   This sequence represents clone C15 from a novel multiple sclerosis
CC   related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC   prophylactic or therapeutic compositions to inhibit expression of a
CC   multiple sclerosis related virus and/or virus associated with rheumatoid
CC   polyarthritis. (Updated on 27-AUG-2003 to correct OS field.) (Updated on
CC   17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1481 BP; 412 A; 410 C; 261 G; 398 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1481;  DB 2;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 3
ADG14846
ID   ADG14846 standard; cDNA; 1481 BP.
XX
AC   ADG14846;
XX
DT   26-FEB-2004  (first entry)
XX
DE   MSRV associated cDNA #9.
XX
KW   ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX
OS   Multiple sclerosis associated retrovirus.
XX
PN   US2003198647-A1.
XX
PD   23-OCT-2003.
XX
PF   03-APR-2002; 2002US-00114104.
XX
PR   26-NOV-1996;   96US-00756429.
PR   26-NOV-1997;   97US-00979847.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2004-032461/03.
XX
PT   New isolated nucleic acid and their fragments having the pol gene of a
PT   retrovirus, useful for diagnosing, preventing and/or treating multiple
PT   sclerosis and/or rheumatoid arthritis.
XX
PS   Disclosure; SEQ ID NO 105; 193pp; English.
XX
CC   The invention relates to an isolated nucleic acid which comprises the pol
CC   gene of a retrovirus associated with multiple sclerosis or rheumatoid
CC   arthritis. The methods and compositions of the present invention are
CC   useful for diagnosing, preventing and/or treating multiple sclerosis
CC   and/or rheumatoid arthritis. The present sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence 1481 BP; 412 A; 410 C; 261 G; 398 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1481;  DB 12;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 4
AAA63826
ID   AAA63826 standard; DNA; 2030 BP.
XX
AC   AAA63826;
XX
DT   06-AUG-2003  (revised)
DT   04-DEC-2000  (first entry)
XX
DE   Nucleotide sequence of the MSRV-1 3' env and LTR regions.
XX
KW   MSRV-1; pol region; long terminal repeat; LTR; RU5 region; retrovirus;
KW   ss.
XX
OS   Multiple sclerosis associated retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1629
FT                   /*tag=  a
FT                   /note= "Contains one termination codon"
FT   sig_peptide     1. .81
FT                   /*tag=  b
FT   CAAT_signal     1800..1807
FT                   /*tag=  c
FT   CAAT_signal     1858..1864
FT                   /*tag=  d
FT   TATA_signal     1906..1911
FT                   /*tag=  e
FT   polyA_signal    1996..2002
FT                   /*tag=  f
XX
PN   WO200047745-A1.
XX
PD   17-AUG-2000.
XX
PF   15-FEB-2000; 2000WO-IB000159.
XX
PR   15-FEB-1999;   99EP-00420041.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Perron H,  Komurian-Pradel F;
XX
DR   WPI; 2000-506097/45.
DR   P-PSDB; AAB08195.
XX
PT   Nucleotide fragment of LTR-RU5 region from Multiple Sclerosis retrovirus
PT   (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological
PT   sample.
XX
PS   Disclosure; Fig 2; 23pp; English.
XX
CC   The present sequence represents the nucleotide sequence corresponding to
CC   the 3' env region and long terminal repeat sequences from clone CL6 of
CC   Multiple Sclerosis retrovirus (MSRV-1). The specification describes a
CC   long terminal repeat (LTR)-RU5 region which encodes the expression of a
CC   MSRV-1 protein. This is unusual for LTRs, in particular in the RU5
CC   region. The sequence includes CAAT and TATA signals which are present in
CC   the U3 and R regions and are not directed towards the CDS indicated in
CC   the features table. Probes and antibodies to the MSRV-1 retrovirus
CC   protein and encoding polynucleotide sequences are used to detect the
CC   presence of MSRV-1 retrovirus in a biological sample. (Updated on 06-AUG-
CC   2003 to correct OS field.)
XX
SQ   Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1481;  DB 3;  Length 2030;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481

RESULT 5
AAA96625
ID   AAA96625 standard; DNA; 1629 BP.
XX
AC   AAA96625;
XX
DT   08-FEB-2001  (first entry)
XX
DE   DNA encoding an envelope (env) protein of MSRV-1.
XX
KW   Envelope protein; MSRV-1; superantigen; autoimmune disease; Vbeta16;
KW   Vbeta17; multiple sclerosis; vaccine; ss.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1629
FT                   /*tag=  a
FT                   /product= "envelope protein"
XX
PN   WO200057185-A1.
XX
PD   28-SEP-2000.
XX
PF   20-MAR-2000; 2000WO-FR000691.
XX
PR   19-MAR-1999;   99FR-00003622.
PR   28-OCT-1999;   99FR-00013755.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Lafont M;
XX
DR   WPI; 2000-638279/61.
DR   P-PSDB; AAB19069.
XX
PT   Detecting superantigen activity, useful for identifying agents for
PT   treatment or prevention of autoimmune disease, from expansion or loss of
PT   particular lymphocyte Vbeta determinants.
XX
PS   Claim 25; Page 123-124; 134pp; French.
XX
CC   The present sequence encodes an envelope protein of MSRV-1. The envelope
CC   protein expressed by the endogenous human retrovirus MRSV-1 has
CC   superantigen activity associated with autoimmune disease. The protien can
CC   be detected using the method of the invention. The specification
CC   describes a process for detecting activity of a superantigen in a
CC   biological sample. The process comprises identifying large scale
CC   expansion or loss of lymphocytes that carry at least one of the Vbeta16
CC   and/or Vbeta17 determinants. The method is used to screen for agents that
CC   inhibit the superantigen, especially those associated with MRSV-1 which
CC   is implicated in autoimmune disease, particularly multiple sclerosis.
CC   These agents are potentially useful for treatment or prevention (e.g. as
CC   vaccines) of autoimmune diseases
XX
SQ   Sequence 1629 BP; 456 A; 454 C; 287 G; 432 T; 0 U; 0 Other;

  Query Match             98.7%;  Score 1461.8;  DB 3;  Length 1629;
  Best Local Similarity   99.2%;  Pred. No. 0;
  Matches 1469;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 120

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              |||||||||||||||||||||  |||||||||||||||||||||||||||| |||||||
Db       1321 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              |||||||||||||||||||||||||| ||| |||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||||||||| || ||  | ||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCA 1481


RESULT 6
ADB84400
ID   ADB84400 standard; DNA; 1481 BP.
XX
AC   ADB84400;
XX
DT   04-DEC-2003  (first entry)
XX
DE   MSRV-1 associated DNA sequence #19.
XX
KW   ds; multiple sclerosis; rheumatoid arthritis; gag; pol;
KW   reverse transcriptase; ribonuclease H.
XX
OS   Unidentified;.
XX
PN   US2003039664-A1.
XX
PD   27-FEB-2003.
XX
PF   26-NOV-1997;   97US-00979847.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (PERR/) PERRON H.
PA   (BESE/) BESEME F.
PA   (BEDI/) BEDIN F.
PA   (PARA/) PARANHOS-BACCALA G.
PA   (KOMU/) KOMURIAN-PRADEL F.
PA   (JOLI/) JOLIVET-REYNAUD C.
PA   (MAND/) MANDRAND B.
PA   (GARS/) GARSON J A.
PA   (TUKE/) TUKE P W.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2003-512253/48.
XX
PT   New isolated or purified nucleic acid associated with multiple sclerosis
PT   and/or rheumatoid arthritis, useful for detecting a virus associated with
PT   multiple sclerosis or rheumatoid arthritis in a biological sample.
XX
PS   Claim 31; Page 77-78; 193pp; English.
XX
CC   The invention relates to an isolated or purified nucleic acid from a
CC   virus associated with multiple sclerosis and/or rheumatoid arthritis,
CC   multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
CC   pol, gag or reverse transcriptase genes (or their fragments) encoding the
CC   proteins or defined peptides (including immunodominant peptides,
CC   antigenic peptides or conserved motifs).Also included are a process for
CC   detecting a virus associated with multiple sclerosis or rheumatoid
CC   arthritis in a biological sample, a nucleic acid probe for the detection
CC   of a virus associated with multiple sclerosis or rheumatoid arthritis, a
CC   primer for the amplification by polymerisation of a nucleic acid of a
CC   viral material associated with multiple sclerosis or rheumatoid
CC   arthritis, a polypeptide exhibiting an inhibitory activity on the
CC   proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC   and an antibody directed against the MSRV-1 virus obtained by
CC   immunologically reacting a human or animal body or cells with an
CC   immunogenic agent consisting of the antigenic polypeptide defined above.
CC   The nucleic acids are useful for detecting a biological sample a virus
CC   associated with multiple sclerosis or rheumatoid arthritis, or for
CC   detecting in a biological sample, the presence of or exposure to a virus
CC   associated with multiple sclerosis or rheumatoid arthritis. The present
CC   sequence is a claimed MSRV-associated sequence whose identity cannot be
CC   accurately determined. Note: The SEQ ID numbers for the sequences as
CC   displayed in the main body of the patent do not match the SEQ ID numbers
CC   in the sequence listing. Consequently those sequences mentioned in the
CC   claims may not be the sequences the authors intended to claim.
XX
SQ   Sequence 1481 BP; 405 A; 403 C; 257 G; 385 T; 0 U; 31 Other;

  Query Match             97.9%;  Score 1450;  DB 9;  Length 1481;
  Best Local Similarity   97.9%;  Pred. No. 0;
  Matches 1450;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGN 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCN 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATN 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTN 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATN 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAN 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACN 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCN 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATN 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTN 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTN 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGN 780
```

```
Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTN 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAN 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATN 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTANN 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACNN 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCNN 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTNN 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGNN 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCNN 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATNN 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAANN 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481

RESULT 7
AAD41225
ID   AAD41225 standard; cDNA; 2074 BP.
XX
AC   AAD41225;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Human EMBRY-2 cDNA.
XX
KW   Human; embryogenesis associated protein; AIDS; reproductive disorder;
KW   infertility; endometriosis; endometrial tumour; inflammatory disorder;
KW   autoimmune disorder; acquired immune deficiency syndrome; transgenic;
KW   ovarian tumour; contact dermatitis; placenta disorder; preeclampsia;
KW   EMBRY-2; allergy; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             64. .1692
FT                   /*tag=  a
FT                   /product= "EMBRY-2 protein"
FT   sig_peptide     64. .123
FT                   /*tag=  b
FT   mat_peptide     124. .1689
FT                   /*tag=  c
FT                   /product= "Mature EMBRY-2 protein"
XX
PN   WO200248362-A2.
XX
PD   20-JUN-2002.
XX
PF   14-NOV-2001; 2001WO-US043956.
XX
PR   15-NOV-2000; 2000US-0249407P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Ramkumar J,  Arvizu C;
XX
DR   WPI; 2002-537629/57.
DR   P-PSDB; AAE25054.
XX
PT   New polypeptides of human embryogenesis associated proteins for screening
PT   modulators useful for treating or preventing disorders e.g.
PT   endometriosis, infertility, allergy, preeclampsia.
XX
PS   Claim 59; Page 96-97; 97pp; English.
XX
CC   The invention relates to human embryogenesis associated proteins (EMBRY)
CC   and nucleic acid molecules encoding such proteins. EMBRY sequences are
CC   useful for screening modulators useful for treating or preventing
CC   disorders associated with abnormal expression of EMBRY. The disorders
CC   treated include reproductive disorders such as infertility,
CC   endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory
CC   disorder such as acquired immune deficiency syndrome (AIDS), allergies,
CC   contact dermatitis; disorders of the placenta such as preeclampsia,
CC   abruptio placentae etc. Sequences of the invention are also useful for
CC   analysing a proteome of a tissue or a cell type. EMBRY proteins are
CC   useful as immunogens for preparing antibodies. Polynucleotides of the
CC   invention are useful for creating knockin humanised animals or transgenic
CC   animals to model human diseases. They are also used in gene therapy. The
CC   present sequence is human EMBRY-2 cDNA
XX
SQ   Sequence 2074 BP; 583 A; 567 C; 392 G; 531 T; 0 U; 1 Other;

  Query Match             93.8%;  Score 1389.8;  DB 6;  Length 2074;
  Best Local Similarity   96.2%;  Pred. No. 0;
  Matches 1424;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              |||||||||||||||  || |||||||||||||||||||| ||| ||||||  |||||||
Db         64 ATGGCCCTCCCTTATTGTATTTTTCTCTTTACTGTTCTCTCACCACCTTTCAGTCTCACT 123

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              |||||| ||||||||  |||||  |||||||||||||||||||||| ||||||||| |||
Db        124 GCACCCTCTCCATGCCACTGTAGGACCAGTAGCTCCCCTTACCAAGCGTTTCTATGGAGA 183

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | ||||| |||  || |||||||||||||||| |||||||||||||||||||||||| ||
Db        184 ATGCGGCGTCCCAGACATATTGATGCCCCATCGTATAGGAGTTTATCTAAGGGAAACCCC 243

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||| ||||||||||||||||||| ||||||||||||||||||||||||||| |||
Db        244 GCCTTCACCGCCCACACCCATATGCCCCACAACTGCTATAACTCTGCCACTCTTTGTATG 303

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 363

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||| |||||||||||||||||||||||||| |||||||||||||||||||| ||
Db        364 GGAGCCACTATCTGTTGGACTTACTTCACCCATACCGGTATGTCTGATGGGGGTGGAGTT 423

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| |||||||||| |||||||||||| ||||||| ||||
Db        424 CAAGATCAGGCAAGAGAAAAACATGTAAAGGAAGTAATCTCCCAACTCACCCGGGTACAT 483

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        484 AGCACCCCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 543

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||| |||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        544 CATACTCACCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 603

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        604 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAGTC 663

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| |||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db        664 CCTGTACCTGAACAATGGAATAACTTCAGCACAGAAATAAACACCACTTCCATTTTAGTA 723

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        724 GGACCTCTTGTTTCCAATCTGGAAATAATCCATACCTCAAACCTCACCTGTGTAAAATTT 783

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||||||  ||||||| |||||| |||||||| |||||||| |||||||||| |
Db        784 AGCAATACTATAGCAACAACCAACTCCCAGTGCATCAGATGGGTAACTCCTCCCACACAA 843

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 903

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 963

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 1023

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1083

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1143

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1203

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1263

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1264 GGAGAAGAATGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1323

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1324 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACACTGGGGCCTC 1383

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1384 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1443

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1444 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1503

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||||||||| |||||  | | |||||||| |||||
Db       1504 GAAGCTGTAAAGCTACAAATGGTTCTTCAAATGGAGCCCCA 1544
```

```
RESULT 8
AAD24195
ID   AAD24195 standard; cDNA; 2930 BP.
XX
AC   AAD24195;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Human syncytin cDNA.
XX
KW   Human; syncytin; preeclampsia; gestational trophoblast disorder;
KW   choriocarcinoma; hydatiform mole; placental site tumour; abortion;
KW   envelope gene; human endogenous defective retrovirus; HERV-W; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             930. .2546
FT                   /*tag=  a
FT                   /product= "Syncytin"
XX
PN   WO200204678-A2.
XX
PD   17-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-US021719.
XX
PR   07-JUL-2000; 2000US-0216657P.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Keith JC,  Mccoy JM,  Mi S;
XX
DR   WPI; 2002-171727/22.
DR   P-PSDB; AAE14540.
XX
PT   Identifying a compound for treating a subject with or at risk of
PT   developing preeclampsia, comprises determining whether the expression or
PT   activity of syncytin in the cell is modulated in the presence of a test
PT   compound.
XX
PS   Disclosure; Page 39-42; 43pp; English.
XX
CC   The invention relates to identifying compounds which are modulators of
CC   syncytin expression. The syncytin modulators are useful in diagnosis and
CC   treatment of preeclampsia and gestational trophoblast disorders (e.g.
CC   choriocarcinoma, hydatiform mole, placental site tumour and missed/
CC   incomplete abortion). Syncytin is a human gene derived from the envelope
CC   gene of human endogenous defective retrovirus, HERV-W. The present
CC   invention is based partly on the discovery that syncytin expression is
CC   dramatically reduced in preeclampsia, and is also mis-localised to the
CC   apical syncytiotrophoblast membrane. The present sequence is human
CC   syncytin cDNA
XX
SQ   Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 U; 0 Other;

  Query Match             90.2%;  Score 1335.4;  DB 6;  Length 2930;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        930 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 989

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        990 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1049

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db       1050 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1109
```

```
Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1110 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1169

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1170 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1229

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1230 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1289

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       1290 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1349

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1350 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1409

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1410 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1469

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1470 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1529

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1589

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1649

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | |||||||||| ||||||||||||||||||||||| |||||||||| |
Db       1650 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1709

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1710 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1769

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1770 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1829

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1830 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1889

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1890 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1949

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1950 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2009

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2010 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2069

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       2070 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2129

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       2130 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2189

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       2190 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2249

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       2250 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2309

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       2310 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2369

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       2370 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2410
```

```
RESULT 9
AAX77526
ID   AAX77526 standard; cDNA; 2946 BP.
XX
AC   AAX77526;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human secreted protein AJ172_2 cDNA.
XX
KW   Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW   bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW   cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW   haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;
KW   haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW   cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
OS   Homo sapiens.
XX
PN   WO9926972-A1.
XX
PD   03-JUN-1999.
XX
PF   17-NOV-1998;   98WO-US024614.
XX
PR   21-NOV-1997;   97US-00976110.
PR   18-MAY-1998;   98US-00080478.
PR   20-OCT-1998;   98US-00175928.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M;
XX
DR   WPI; 1999-357813/30.
DR   P-PSDB; AAY08622.
XX
PT   New polynucleotides encoding secreted proteins.
XX
PS   Claim 13a; Page 100-101; 142pp; English.
XX
CC   This invention describes novel human secreted proteins encoded by
CC   polynucleotides isolated from human adult testes, adult brain, adult
CC   blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC   libraries. The products of the invention are predicted to have biological
CC   activities which would make them suitable for treating, preventing or
CC   ameliorating medical conditions in humans and animals, although no
CC   supporting data is given. Suggested activities include nutritional
CC   activity, cytokine and cell proliferation/differentiation activity,
CC   immune stimulating (e.g. as vaccines) or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, and tumour
CC   inhibition activity. The polynucleotides are also stated to be useful for
CC   gene therapy
```

```
XX
SQ   Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

  Query Match             90.2%;  Score 1335.4;  DB 2;  Length 2946;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        928 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 987

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        988 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1047

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| |||||||||| | |||||||||| | ||
Db       1048 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1107

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1108 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1167

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1168 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1227

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1228 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1287

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       1288 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1347

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1348 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1407

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1408 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1467

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1468 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1527

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1528 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1587

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1588 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1647

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       1648 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1707

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1708 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1767

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1768 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1827

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1828 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1887

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1888 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1947

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1948 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2007

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2008 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2067

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       2068 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2127

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       2128 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2187

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       2188 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2247

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       2248 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2307

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       2308 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2367

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       2368 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2408


RESULT 10
AAZ59468
ID   AAZ59468 standard; cDNA; 2946 BP.
XX
AC   AAZ59468;
XX
DT   11-APR-2000  (first entry)
XX
DE   Human secreted protein AJ172_2 polynucleotide sequence.
XX
KW   Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW   placental pathology; metastasis inhibition; nutritional activity;
KW   immune stimulator; haematopoiesis regulator; tissue growth;
KW   tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9960020-A1.
XX
PD   25-NOV-1999.
XX
PF   17-MAY-1999;   99WO-US010915.
XX
PR   18-MAY-1998;   98US-00080478.
PR   20-OCT-1998;   98US-00175928.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Mi S,  Treacy M;
XX
DR   WPI; 2000-116311/10.
DR   P-PSDB; AAY67313.
XX
PT   New polynucleotides encoding secreted cDNA libraries, used to develop
PT   products for the diagnosis and treatment of neoplastic disease.
XX
```

```
PS   Claim 14; Page 107-108; 149pp; English.
XX
CC   This is the human secreted protein AJ172_2 nucleotide sequence, obtained
CC   from a human adult testes cDNA library. The invention relates to secreted
CC   human and murine proteins. The polynucleotides and proteins are predicted
CC   to have biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals. Detection of the levels of the proteins can be used for the
CC   diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
CC   which modulate the expression or function of the proteins may be used for
CC   treating a neoplastic disease and inhibiting metastasis. Other suggested
CC   activities include nutritional activity (e.g. in feeds), cytokine and
CC   cell proliferation/differentiation activity, immune stimulating (e.g. as
CC   vaccines) or suppressing activity, haematopoiesis regulating activity,
CC   tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC   invasion suppressor activity, and tumour inhibition activity. The
CC   polynucleotide sequences are also stated to be useful for gene therapy
XX
SQ   Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

  Query Match             90.2%;  Score 1335.4;  DB 3;  Length 2946;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db         928 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 987

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         988 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1047

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        1048 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1107

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        1108 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1167

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        1168 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1227

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        1228 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1287

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        1288 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1347

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        1348 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1407

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        1408 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1467

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        1468 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1527

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1528 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1587

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1588 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1647

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db        1648 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1707

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        1708 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1767

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1768 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1827

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db        1828 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1887

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        1888 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1947

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1948 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2007

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2008 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2067

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        2068 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2127

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db        2128 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2187

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db        2188 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2247

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        2248 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2307

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db        2308 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2367

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||| ||||| ||||      | ||| |  || |
Db        2368 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2408


RESULT 11
ADC38776
ID   ADC38776 standard; cDNA; 2946 BP.
XX
AC   ADC38776;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human cDNA encoding a secreted protein #63.
XX
KW   ss; gene; immune disorder; severe combined immunodeficiency; SCID;
KW   autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW   rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;
KW   lymphoid cell deficiency; osteoporosis; osteoarthritis;
KW   peripheral nervous system disease; peripheral neuropathy;
KW   Alzheimer's disease; Parkinson's disease; coagulation disorder;
KW   inflammatory disease; systemic inflammatory response syndrome; SIRS;
KW   ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;
KW   hypersensitivity; regeneration; neural cell proliferation; fertility;
KW   tumour; chemokine; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   US2002193567-A1.
XX
PD   19-DEC-2002.
XX
PF   02-APR-2002; 2002US-00114893.
XX
PR   11-AUG-1995;   95US-00514014.
PR   05-APR-1996;   96US-00628364.
PR   19-APR-1996;   96US-00635311.
PR   07-JUN-1996;   96US-00659224.
PR   17-JUN-1996;   96US-00664596.
PR   09-JUL-1996;   96US-00677231.
PR   26-JUL-1996;   96US-00686878.
PR   23-AUG-1996;   96US-00701819.
PR   27-SEP-1996;   96US-00721488.
PR   27-SEP-1996;   96US-00721798.
PR   27-SEP-1996;   96US-00721923.
PR   27-SEP-1996;   96US-00721926.
PR   25-OCT-1996;   96US-00738367.
PR   30-OCT-1996;   96US-00739775.
PR   13-JAN-1997;   97US-00783395.
PR   10-APR-1997;   97US-00833823.
PR   02-JUN-1997;   97US-00867677.
PR   05-SEP-1997;   97US-00924838.
PR   06-OCT-1999;   99US-00413232.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR,  Spaulding V,  Carlin-Duckett M;
PI   Kelleher K;
XX
DR   WPI; 2003-657236/62.
DR   P-PSDB; ADC38777.
XX
PT   Proteins AZ3021 encoded by clone AZ3021 from human adult colon, and
PT   BD12716 encoded by clone BD12716 from human fetal kidney cDNA library,
PT   useful for treating e.g. multiple sclerosis and rheumatoid arthritis.
XX
PS   Disclosure; SEQ ID NO 134; 412pp; English.
XX
CC   The invention relates to a protein comprising fully defined AZ302 1
CC   protein or BD127 1 6 protein. The polynucleotides are useful for
CC   expressing recombinant proteins for analysis and are also useful as
CC   chromosome markers or tags to identify chromosomes or to map related gene
CC   positions. The proteins are useful as amino acid supplement, carbon
CC   source, nitrogen source and carbohydrate source. The proteins are useful
CC   for treating various immune deficiencies and disorders (e.g. severe
CC   combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC   sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic
CC   reactions (e.g. asthma), myeloid or lymphoid cell deficiencies,
CC   osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g.
CC   peripheral neuropathy, Alzheimer's disease, Parkinson's disease),
CC   coagulation disorders, inflammatory diseases (e.g. systemic inflammatory
CC   response syndrome (SIRS), ischaemia-reperfusion injury, Crohn's disease),
CC   anaphylaxis and hypersensitivity. Proteins are also useful for inducing
CC   tumour immunity, for inducing bone, cartilage, tendon, ligament and/or
CC   nerve growth or regeneration, for proliferating neural cells and for
CC   regenerating nerve and brain tissue, for inducing fertility and for
CC   inhibiting tumour growth. Proteins are also useful as chemokine for
CC   mammalian cells (e.g., monocytes, fibroblasts, neutrophils), and also
CC   useful as inhibitors of receptor/ligand interactions. The present
CC   sequence represents cDNA encoding a human secreted protein.
XX
SQ   Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

  Query Match             90.2%;  Score 1335.4;  DB 10;  Length 2946;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        928 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 987

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        988 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1047

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db       1048 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1107

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1108 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1167

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1168 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1227

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1228 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1287

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       1288 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1347

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1348 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1407

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1408 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1467

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1468 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1527

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1528 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 1587

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1588 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1647

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | |||||||||| ||||||||||||||||||||||| |||||||||| |
Db       1648 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1707

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1708 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1767

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1768 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1827

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db       1828 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1887

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1888 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1947

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1948 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2007
```

```
Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2008 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2067

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       2068 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2127

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       2128 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2187

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       2188 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2247

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       2248 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2307

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db       2308 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2367

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       2368 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2408


RESULT 12
AAH20070
ID   AAH20070 standard; DNA; 1617 BP.
XX
AC   AAH20070;
XX
DT   08-AUG-2001  (first entry)
XX
DE   HERV-W envelope protein G coding sequence.
XX
KW   Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW   envelope protein; multiple sclerosis-related superantigen; vaccine;
KW   surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW   antisense-therapy; autoimmune disorder; ds.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1617
FT                   /*tag=  a
FT                   /product= "HERV-W envelope protein G"
XX
PN   WO200131021-A1.
XX
PD   03-MAY-2001.
XX
PF   30-OCT-2000; 2000WO-EP010659.
XX
PR   28-OCT-1999;   99EP-00402690.
XX
PA   (UYGE-) UNIV GENEVE.
XX
PI   Conrad B,  Mach B;
XX
DR   WPI; 2001-316336/33.
DR   P-PSDB; AAB75138.
XX
PT   New human retrovirus HERV-W ENV proteins/peptides having superantigen
PT   activity useful for diagnosing and treating multiple sclerosis.
XX
PS   Claim 13; Fig 10; 94pp; English.
XX
CC   On the basis of the PBS t-RNA motif used for the classification of human
CC   endogenous retrovirus (HERVs) the full length endogenous provirus which
CC   was been located on the long arm of human chromosome 7 (7q21-22) has been
CC   designated HERV-W. The present invention describes proteins or peptides
CC   (I) having superantigen (SAg) activity comprising the ENV protein (ENV)
CC   of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC   have neuroprotective activity, and can be used in: vaccines; antisense-
CC   therapy; and HERV-W SAg activity-inhibitors. (I) and encoding DNA/RNA are
CC   useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC   disorders. (I) are also useful for identifying substances (and optionally
CC   recovering) capable of binding to a retroviral superantigen associated
CC   with MS, substances capable of blocking SAg activity and substances
CC   capable of blocking transcription or translation of HERV-W retroviral
CC   superantigen. A protein or peptide derived from (I), modified to be
CC   devoid of SAg activity and being capable of generating an immune response
CC   against HERV-W retroviral SAg is useful in therapy. Nucleic acid
CC   molecules encoding (I) are useful as vaccines against MS. Substances
CC   capable of blocking SAg activity, capable of binding to a retroviral
CC   superantigen associated with MS, or capable of blocking transcription or
CC   translation of HERV-W retroviral superantigen for use in treating or
CC   preventing MS, obtained using (I) are useful for the treatment and
CC   prevention of MS. (I) and nucleic acids encoding them are useful for
CC   diagnosing autoimmune disease. The present sequence encodes the
CC   specifically claimed envelope protein of HERV-W designated G
XX
SQ   Sequence 1617 BP; 442 A; 452 C; 296 G; 427 T; 0 U; 0 Other;

  Query Match            90.1%;  Score 1333.8;  DB 5;  Length 1617;
  Best Local Similarity  93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db          1 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         61 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        121 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        301 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        361 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        421 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        481 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        541 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
```

```
Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db        721 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        961 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       1201 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       1261 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1381 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       1441 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1481


RESULT 13
ABN97948
ID   ABN97948 standard; DNA; 2055 BP.
XX
AC   ABN97948;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral HERV-7q env coding sequence.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 3; Page 142-145; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 2055 BP; 576 A; 574 C; 376 G; 529 T; 0 U; 0 Other;

  Query Match             90.1%;  Score 1333.8;  DB 3;  Length 2055;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        391 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 450

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        451 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 510

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        511 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 570

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        571 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 630

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        631 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 690

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        691 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 750

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        751 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 810

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        811 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 870

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db     871 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 930

Qy     541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
           ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db     931 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 990

Qy     601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     991 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1050

Qy     661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1051 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1110

Qy     721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
           |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db    1111 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1170

Qy     781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    1171 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1230

Qy     841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1231 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1290

Qy     901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
           ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db    1291 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1350

Qy     961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
           ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db    1351 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1410

Qy    1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
           ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db    1411 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1470

Qy    1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
           ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1471 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1530

Qy    1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
           ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db    1531 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1590

Qy    1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
           || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db    1591 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1650

Qy    1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
           |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db    1651 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1710

Qy    1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
           ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db    1711 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1770

Qy    1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
            ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||| 
Db    1771 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1830

Qy    1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
           ||||||||||| ||||| ||||      | ||| |  || |
Db    1831 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1871

RESULT 14
ABN97927
ID   ABN97927 standard; DNA; 2599 BP.
XX
AC   ABN97927;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral sequence HERV 7 env.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 1; Page 128-129; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 2599 BP; 744 A; 718 C; 495 G; 642 T; 0 U; 0 Other;

  Query Match             90.1%;  Score 1333.8;  DB 3;  Length 2599;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy       1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
           ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db     929 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 988

Qy      61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
           ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db     989 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1048

Qy     121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
           | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db    1049 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1108

Qy     181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    1109 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1168

Qy     241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    1169 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1228

Qy     301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
           |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db    1229 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1288
```

```
Qy    361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
          |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db   1289 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1348

Qy    421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
           ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   1349 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1408

Qy    481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
          |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db   1409 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1468

Qy    541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
          ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db   1469 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1528

Qy    601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1529 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1588

Qy    661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1589 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1648

Qy    721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
          |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db   1649 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1708

Qy    781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   1709 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1768

Qy    841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1769 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1828

Qy    901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
          ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db   1829 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1888

Qy    961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
          ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db   1889 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1948

Qy   1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
          ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db   1949 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2008

Qy   1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
          ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2009 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2068

Qy   1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
          ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db   2069 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2128

Qy   1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
          || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db   2129 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2188

Qy   1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
          |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db   2189 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2248

Qy   1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
          ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db   2249 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2308

Qy   1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db   2309 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2368

Qy   1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
```

```
          ||||||||||| ||||| ||||      | ||| |  || |
Db   2369 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2409

RESULT 15
ABN97929
ID   ABN97929 standard; DNA; 10499 BP.
XX
AC   ABN97929;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral sequence HERV-7q.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 3; Fig 1; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 U; 0 Other;

  Query Match             90.1%;  Score 1333.8;  DB 3;  Length 10499;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy      1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
          ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db   7879 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 7938

Qy     61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
          ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db   7939 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 7998

Qy    121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
          | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db   7999 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 8058

Qy    181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db       8059 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 8118

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       8119 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 8178

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       8179 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 8238

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       8239 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 8298

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       8299 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 8358

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       8359 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 8418

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       8419 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 8478

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8479 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 8538

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8539 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 8598

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | |||||||||| ||||||||||||||||||||||| |||||||||| |
Db       8599 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 8658

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       8659 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 8718

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8719 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 8778

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db       8779 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 8838

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       8839 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 8898

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       8899 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 8958

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8959 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 9018

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       9019 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 9078

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       9079 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 9138

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       9139 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 9198

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       9199 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 9258

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db       9259 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 9318

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       9319 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 9359
```

```
RESULT 16
ABL61744
ID   ABL61744 standard; DNA; 56093 BP.
XX
AC   ABL61744;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Colon adenocarcinoma related gene sequence SEQ ID NO:81.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 81; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 U; 0 Other;

  Query Match             90.1%;  Score 1333.8;  DB 6;  Length 56093;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db      35879 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 35938

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db      35939 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 35998

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| || | |||||||||| | ||
Db      35999 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 36058

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      36059 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 36118

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db      36119 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 36178

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  ||||||||||||||||||||| ||
Db      36179 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 36238

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| ||||||||||||| |||||||| ||||
Db      36239 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 36298

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      36299 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 36358

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db      36359 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 36418

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| ||||||||||||| |||||||||||||  ||||||||
Db      36419 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 36478

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36479 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 36538

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36539 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 36598

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||||| |||||||||||||||||||||||| |||||||||| |
Db      36599 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 36658

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      36659 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 36718

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36719 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 36778

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| ||||||||||||||||||||
Db      36779 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 36838

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||||   |||||||||||||||||| |||||
Db      36839 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 36898

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||||| |
Db      36899 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 36958

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36959 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 37018

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db      37019 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 37078

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| |||||||||||||||||||||||| |||| ||||||||||||||||||||
Db      37079 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 37138

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db      37139 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 37198

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              |||||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db      37199 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 37258

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db      37259 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 37318

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db      37319 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 37359

RESULT 17
```

```
AAF55630
ID   AAF55630 standard; DNA; 2781 BP.
XX
AC   AAF55630;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
KW   Envelope protein; HERV; syncytia formation; placental development;
KW   syncytia; cancer; cell adhesion; ss.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             762. .2378
FT                   /*tag=  a
FT                   /product= "envelope protein"
XX
PN   WO200116171-A1.
XX
PD   08-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-FR002429.
XX
PR   01-SEP-1999;   99FR-00011141.
PR   15-SEP-1999;   99FR-00011793.
XX
PA   (INMR ) BIO MERIEUX.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI   Mallet F,  Cosset F,  Blond J,  Lavillette D,  Bouton O,  Ruggieri A;
XX
DR   WPI; 2001-226676/23.
DR   P-PSDB; AAB67652.
XX
PT   Detecting expression of human endogenous retrovirus envelope protein in
PT   cells of a tissue or culture, from its ability to induce syncytia.
XX
PS   Disclosure; Page 44-45; 57pp; French.
XX
CC   The present sequence encodes a human endogenous retrovirus envelope
CC   protein. The specification describes a method for detecting expression of
CC   an envelope protein from a human endogenous retrovirus (HERV), in cells,
CC   of a tissue or culture. The method comprises detecting syncytia formation
CC   due to the fusogenic properties of the envelope protein. Envelope
CC   polypeptides and polynucleotides are used to produce therapeutic or
CC   prophylactic compositions, particularly for treatment of cancer, to
CC   correct defects in placental development (or other natural formation of
CC   other types of syncytia), and to promote adhesion of cells in grafts or
CC   cellular repair processes. Expression of sequences antisense to the
CC   polynucleotide are used to prevent formation of syncytia
XX
SQ   Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 U; 0 Other;

  Query Match             89.8%;  Score 1330.6;  DB 5;  Length 2781;
  Best Local Similarity   93.7%;  Pred. No. 0;
  Matches 1387;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        762 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 821

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        822 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 881

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        882 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 941

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        942 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1001

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1002 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1061

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1062 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1121

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       1122 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1181

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1182 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1241

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1242 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1301

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1302 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1361

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1362 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1421

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1422 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1481

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       1482 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1541

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1542 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1601

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1602 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1661

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db       1662 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1721

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  |||| |||||||||||   ||||||||||||||||||| |||||
Db       1722 CTTCCTTTTGTTATAGGAGCGGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1781

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1782 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1841

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | |||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db       1842 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAGCTTAACTCCCTAGCAGCAGTAGTCCTT 1901

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1902 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1961

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       1962 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2021

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       2022 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2081
```

```
Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       2082 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2141

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db       2142 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2201

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | |||  |  || |
Db       2202 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2242


RESULT 18
AAA59215
ID   AAA59215 standard; DNA; 7582 BP.
XX
AC   AAA59215;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Human endogenous retrovirus W (HERV-W) sequence.
XX
KW   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW   gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   LTR             1. .120
FT                   /*tag=  a
FT                   /note= "R of 5' LTR"
FT   LTR             121. .575
FT                   /*tag=  b
FT                   /note= "U5 of 5' LTR"
FT   primer_bind     579. .596
FT                   /*tag=  c
FT   CDS             5581. .7194
FT                   /*tag=  d
FT                   /note= "ORF1 env538"
FT   CDS             7039. .7194
FT                   /*tag=  e
FT                   /note= "ORF2 52 AA"
FT   CDS             7112. .7255
FT                   /*tag=  f
FT                   /note= "ORF3 48 AA"
FT   misc_feature    7244. .7254
FT                   /*tag=  g
FT                   /note= "polypurine tract"
FT   LTR             7256. .7582
FT                   /*tag=  h
FT                   /note= "U3-R of 3' LTR"
FT   polyA_signal    7563. .7569
FT                   /*tag=  i
XX
PN   WO200043521-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-FR000144.
XX
PR   21-JAN-1999;   99FR-00000888.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Mallet F,  Voisset C;
XX
DR   WPI; 2000-499229/44.
XX
PT   New nucleic acid from human endogenous retrovirus, useful e.g. for
PT   diagnosis of autoimmune disease and complications of pregnancy, contains
PT   at least part of the gag gene.
XX
PS   Disclosure; Page 49-52; 53pp; French.
XX
CC   The present sequence represents an endogenetic retrovirus, which is
CC   associated with an autoimmune disease, and is integrated into the human
CC   genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
CC   HERV-W retrovirus is associated with autoimmune disease, failure of
CC   pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
CC   proteins derived from it, are useful for diagnosis of autoimmune disease
CC   (specifically multiple sclerosis) and for monitoring pregnancy. The
CC   nucleic acid fragments may also be used for in situ labelling of isolated
CC   chromosomes, while the transcription product can be used to study or
CC   monitor T cell proliferation in vitro
XX
SQ   Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 0 U; 216 Other;

  Query Match             89.4%;  Score 1324.2;  DB 3;  Length 7582;
  Best Local Similarity   92.2%;  Pred. No. 0;
  Matches 1365;  Conservative   24;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||:|||||||||||||| ||||||||:||:||||:| |:|||| :|||| ||||||||
Db       5581 ATGGSCCTCCCTTATCATATTTTTCTCTKTASTGTTSTTTYACCCTSTTTCACTCTCACT 5640

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||:|||:|||||||||||| |||
Db       5641 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTYACCMAGAGTTTCTATGGAGA 5700

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |:|||||||| | ||
Db       5701 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTSTAAGGGAACCCCC 5760

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       5761 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 5820

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       5821 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 5880

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       5881 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 5940

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| |||:||| ||||
Db       5941 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCSGGGTACAT 6000

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       6001 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 6060

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       6061 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 6120

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| ||||||:||||||  ||||||||
Db       6121 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCARGCCATATGTTTCAATC 6180

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6181 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 6240

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       6241 GGACCTCTTGTTTCCAATSTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 6300

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       6301 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 6360

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
```

```
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        6361 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 6420

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               |||||||||||||||||||||||||||||||||||||||||||||||:|||:||||||||
Db        6421 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCYATGRCCATCTAC 6480

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db        6481 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 6540

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        6541 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 6600

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        6601 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 6660

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6661 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 6720

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               |:||||||||||||||||||||:|||||||||  |:||||||||||||||||||||||||
Db        6721 CRAAATCGAAGAGCTTTAGACTYGCTAACCGCTGARAGAGGGGGAACCTGTTTATTTTTA 6780

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| ||||||||||||||||||||||| |||| |||||||||||||||:|||||
Db        6781 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTRAAGAA 6840

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               ||||:||||||||||||| |||:||||||:||||||| |||||| | || ||||||||||
Db        6841 ATTCSAGATCGAATACAACGTAKAGCAGARGAGCTTCGAAACACTGGACCCTGGGGCCTC 6900

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               |||||||:||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        6901 CTCAGCCRATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 6960

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                ||||||||||||||||||||||||||:||||||||||||| |||||||||||||||||
Db        6961 CTACTCCTCTTTGGACCCTGTATCTTTRACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 7020

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||:|| ||||| ||||      | ||| |  || |
Db        7021 GAAGCTGTRAAACTACAAATGGAGCCCAAGATGCAGTCCAA 7061

RESULT 19
AAX25665
ID   AAX25665 standard; cDNA to mRNA; 7582 BP.
XX
AC   AAX25665;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Complete human endogenous retrovirus W genome.
XX
KW   Clone; human endogenous retrovirus; genome; autoimmune disease;
KW   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW   disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9902696-A1.
XX
PD   21-JAN-1999.
XX
PF   06-JUL-1998;    98WO-FR001442.
XX
PR   07-JUL-1997;    97FR-00008815.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Beseme F,  Blond J,  Bouton O,  Mandrand B,  Mallet F;
XX
DR   WPI; 1999-120897/10.
XX
PT   New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT   exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT   disease, and abnormal or failed pregnancy.
XX
PS   Claim 1; Page 71-74; 106pp; French.
XX
CC   This sequence represents the complete sequence of the human endogenous
CC   retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC   peptides encoded by them are markers of autoimmune disease (e.g. multiple
CC   sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC   insulin- dependent diabetes and related pathologies) and of abnormal or
CC   unsuccessful pregnancy and can be used as chromosomal markers for
CC   susceptibility to these conditions, or proximity markers of genes
CC   associated with this susceptibility
XX
SQ   Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 Other;

  Query Match             89.3%;  Score 1322.6;  DB 2;  Length 7582;
  Best Local Similarity   92.1%;  Pred. No. 0;
  Matches 1364;  Conservative   24;  Mismatches   93;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||:|||||||||||||| ||||||||:||:||||:| |:|||| :|||| ||||||||
Db        5581 ATGGSCCTCCCTTATCATATTTTTCTCTKTASTGTTSTTTYACCCTSTTTCACTCTCACT 5640

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||:|||:|||||||||||| |||
Db        5641 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTYACCMAGAGTTTCTATGGAGA 5700

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |:|||||||| | ||
Db        5701 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTSTAAGGGAACCCCC 5760

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        5761 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 5820

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        5821 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 5880

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        5881 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 5940

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| |||:||| ||||
Db        5941 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCSGGGTACAT 6000

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        6001 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 6060

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        6061 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 6120

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||| ||||||:||||||  ||||||||
Db        6121 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCARGCCATATGTTTCAATC 6180

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6181 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 6240

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        6241 GGACCTCTTGTTTCCAATSTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 6300
```

```
Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       6301 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 6360

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       6361 ATAGTCTGCCTACCCTCAGGAATATTTTTTCTCTGTGGTACCTCAGCCTATCGTTGTTTG 6420

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              |||||||||||||||||||||||||||||||||||||||||||||||:|||:||||||||
Db       6421 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCYATGRCCATCTAC 6480

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db       6481 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 6540

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       6541 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 6600

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       6601 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 6660

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6661 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 6720

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              |:||||||||||||||||||||:|||||||||  |:||||||||||||||||||||||||
Db       6721 CRAAATCGAAGAGCTTTAGACTYGCTAACCGCTGARAGAGGGGGAACCTGTTTATTTTTA 6780

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| |||||||||||||||||||||||| |||| ||||||||||||||:|||||
Db       6781 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTRAAGAA 6840

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||:||||||||||||| |||:||||||:||||||| |||||| | || ||||||||||
Db       6841 ATTCSAGATCGAATACAACGTAKAGCAGARGAGCTTCGAAACACTGGACCCTGGGGCCTC 6900

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              |||||||:||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       6901 CTCAGCCRATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 6960

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               ||||||||||||||||||||||||||:||||||||||||| |||||||||||||||||
Db       6961 CTACTCCTCTTTGGACCCTGTATCTTTRACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 7020

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||:|| ||||| ||||      | ||| |  || |
Db       7021 GAAGCTGTRAAACTACAAATGGAGCCCAAGATGCAGTCCAA 7061

RESULT 20
AAH20069
ID   AAH20069 standard; DNA; 2782 BP.
XX
AC   AAH20069;
XX
DT   08-AUG-2001  (first entry)
XX
DE   HERV-W envelope protein G encoding nucleic acid.
XX
KW   Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW   envelope protein; multiple sclerosis-related superantigen; vaccine;
KW   surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW   antisense-therapy; autoimmune disorder; ds.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1. .762
FT                   /*tag=  a
FT   CDS             763. .2379
FT                   /*tag=  b
FT                   /product= "HERV-W envelope protein G"
FT                   /transl_except= (pos:790. .792,aa:Phe)
FT                   /transl_except= (pos:793. .795,aa:Thr)
FT                   /transl_except= (pos:812. .814,aa:Leu)
FT                   /transl_except= (pos:818. .820,aa:Ser)
FT                   /transl_except= (pos:862. .864,aa:Tyr)
FT                   /transl_except= (pos:865. .867,aa:Gln)
FT                   /transl_except= (pos:1174. .1176,aa:Arg)
FT                   /transl_except= (pos:1441. .1443,aa:Leu)
FT                   /transl_except= (pos:1903. .1905,aa:Gln)
FT                   /transl_except= (pos:2017. .2019,aa:Lys)
FT                   /transl_except= (pos:2026. .2028,aa:Arg)
FT                   /transl_except= (pos:2044. .2046,aa:Arg)
FT                   /transl_except= (pos:2089. .2091,aa:Gln)
FT                   /transl_except= (pos:2170. .2172,aa:Asn)
FT   3'UTR           2380. .2782
FT                   /*tag=  c
XX
PN   WO200131021-A1.
XX
PD   03-MAY-2001.
XX
PF   30-OCT-2000; 2000WO-EP010659.
XX
PR   28-OCT-1999;   99EP-00402690.
XX
PA   (UYGE-) UNIV GENEVE.
XX
PI   Conrad B,  Mach B;
XX
DR   WPI; 2001-316336/33.
DR   P-PSDB; AAB75138.
XX
PT   New human retrovirus HERV-W ENV proteins/peptides having superantigen
PT   activity useful for diagnosing and treating multiple sclerosis.
XX
PS   Claim 13; Fig 9; 94pp; English.
XX
CC   On the basis of the PBS t-RNA motif used for the classification of human
CC   endogenous retrovirus (HERVs) the full length endogenous provirus which
CC   was been located on the long arm of human chromosome 7 (7q21-22) has been
CC   designated HERV-W. The present invention describes proteins or peptides
CC   (I) having superantigen (SAg) activity comprising the ENV protein (ENV)
CC   of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC   have neuroprotective activity, and can be used in: vaccines; antisense-
CC   therapy; and HERV-W SAg activity-inhibitors. (I) and encoding DNA/RNA are
CC   useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC   disorders. (I) are also useful for identifying substances (and optionally
CC   recovering) capable of binding to a retroviral superantigen associated
CC   with MS, substances capable of blocking SAg activity and substances
CC   capable of blocking transcription or translation of HERV-W retroviral
CC   superantigen. A protein or peptide derived from (I), modified to be
CC   devoid of SAg activity and being capable of generating an immune response
CC   against HERV-W retroviral SAg is useful in therapy. Nucleic acid
CC   molecules encoding (I) are useful as vaccines against MS. Substances
CC   capable of blocking SAg activity, capable of binding to a retroviral
CC   superantigen associated with MS, or capable of blocking transcription or
CC   translation of HERV-W retroviral superantigen for use in treating or
CC   preventing MS, obtained using (I) are useful for the treatment and
CC   prevention of MS. (I) and nucleic acids encoding them are useful for
CC   diagnosing autoimmune disease. The present sequence encodes the
CC   specifically claimed envelope protein of HERV-W designated G
XX
SQ   Sequence 2782 BP; 741 A; 768 C; 564 G; 709 T; 0 U; 0 Other;

  Query Match             88.3%;  Score 1308.2;  DB 5;  Length 2782;
  Best Local Similarity   92.7%;  Pred. No. 0;
  Matches 1373;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
```

```
               ||||||||||||||||||| |||||||| || |||||| | ||||  |||| ||||||||
Db         763 ATGGCCCTCCCTTATCATATTTTTCTCTGTAGTGTTCTTTCACCCTGTTTCACTCTCACT 822

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  ||||||||||||||| ||| |||||||||||| |||
Db         823 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTCACCCAGAGTTTCTATGGAGA 882

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db         883 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 942

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         943 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1002

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        1003 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1062

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        1063 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1122

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||| ||| ||||
Db        1123 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCGGGGTACAT 1182

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        1183 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1242

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        1243 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1302

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        1303 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1362

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1363 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1422

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        1423 GGACCTCTTGTTTCCAATGTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1482

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db        1483 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1542

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        1543 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1602

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1603 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1662

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db        1663 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1722

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        1723 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1782

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1783 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1842

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1843 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1902

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               | ||||||||||||||||||||||||||||||  | ||||||||||||||||||||||||
Db        1903 CGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTTA 1962

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| ||||||||||||||||||||||| |||| ||||||||||||||| |||||
Db        1963 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGAA 2022

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||| ||||||||||||| |||| ||||||||||||| |||||| | || ||||||||||
Db        2023 ATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2082

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||| ||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        2083 CTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2142

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||| ||||||||||||| ||||||||||||||||| 
Db        2143 CTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2202

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               |||||||| || ||||| ||||      | ||| |  || |
Db        2203 GAAGCTGTGAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2243


RESULT 21
AAX25661
ID   AAX25661 standard; cDNA to mRNA; 2782 BP.
XX
AC   AAX25661;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Human endogenous retrovirus W clone cl.PH74.
XX
KW   Clone; human endogenous retrovirus; genome; autoimmune disease;
KW   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW   disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9902696-A1.
XX
PD   21-JAN-1999.
XX
PF   06-JUL-1998;   98WO-FR001442.
XX
PR   07-JUL-1997;   97FR-00008815.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Beseme F,  Blond J,  Bouton O,  Mandrand B,  Mallet F;
XX
DR   WPI; 1999-120897/10.
XX
PT   New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT   exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT   disease, and abnormal or failed pregnancy.
XX
PS   Claim 1; Page 60-63; 106pp; French.
XX
CC   This sequence represents clone cl.PH74 of the human endogenous retrovirus
CC   (HERV) W genome. The nucleic acids, their fragments or peptides encoded
CC   by them are markers of autoimmune disease (e.g. multiple sclerosis,
CC   rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC   dependent diabetes and related pathologies) and of abnormal or
CC   unsuccessful pregnancy and can be used as chromosomal markers for
CC   susceptibility to these conditions, or proximity markers of genes
CC   associated with this susceptibility
XX
SQ   Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;
```

```
  Query Match             88.2%;  Score 1306.6;  DB 2;  Length 2782;
  Best Local Similarity   92.6%;  Pred. No. 0;
  Matches 1372;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              |||| |||||||||||||| |||||||| || |||||| | ||||  |||| ||||||||
Db        763 ATGGGCCTCCCTTATCATATTTTTCTCTGTAGTGTTCTTTCACCCTGTTTCACTCTCACT 822

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  ||||||||||||||| ||| |||||||||||| |||
Db        823 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTCACCCAGAGTTTCTATGGAGA 882

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        883 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 942

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        943 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1002

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1003 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1062

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1063 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1122

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||| ||| ||||
Db       1123 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCGGGGTACAT 1182

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1183 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1242

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1243 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1302

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1303 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1362

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1422

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1423 GGACCTCTTGTTTCCAATGTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1482

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | |||||||||| ||||||||||||||||||||||| |||||||||| |
Db       1483 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1542

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1543 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1602

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1662

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db       1663 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1722

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1723 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1782

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1783 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1842

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1902

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              | ||||||||||||||||||||||||||||||  | ||||||||||||||||||||||||
Db       1903 CGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTTA 1962

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| ||||||||||||||| |||||
Db       1963 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGAA 2022

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||| ||||||||||||| |||| ||||||||||||| |||||| | || ||||||||||
Db       2023 ATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2082

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||| ||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       2083 CTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2142

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||| ||||||||||||| |||||||||||||||||
Db       2143 CTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2202

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||| || ||||| ||||      | ||| |  || |
Db       2203 GAAGCTGTGAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2243


RESULT 22
AAA59211
ID   AAA59211 standard; DNA; 2782 BP.
XX
AC   AAA59211;
XX
DT   07-NOV-2000  (first entry)
XX
DE   5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.
XX
KW   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW   gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200043521-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-FR000144.
XX
PR   21-JAN-1999;   99FR-00000888.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Mallet F,  Voisset C;
XX
DR   WPI; 2000-499229/44.
XX
PT   New nucleic acid from human endogenous retrovirus, useful e.g. for
PT   diagnosis of autoimmune disease and complications of pregnancy, contains
PT   at least part of the gag gene.
XX
PS   Disclosure; Page 46-47; 53pp; French.
XX
CC   The present sequence represents an endogenetic retroviral nucleic acid
CC   fragment, which is associated with an autoimmune disease, and is
CC   integrated into the human genome. The fragment is originally derived from
CC   a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W
CC   retrovirus is associated with autoimmune disease, failure of pregnancy or
CC   disorders of pregnancy. The nucleic acid fragment, or proteins derived
CC   from it, are useful for diagnosis of autoimmune disease (specifically
CC   multiple sclerosis) and for monitoring pregnancy. The nucleic acid
CC   fragments may also be used for in situ labelling of isolated chromosomes,
CC   while the transcription product can be used to study or monitor T cell
CC   proliferation in vitro
XX
SQ   Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;

  Query Match             88.2%;  Score 1306.6;  DB 3;  Length 2782;
  Best Local Similarity   92.6%;  Pred. No. 0;
  Matches 1372;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              |||| |||||||||||||| |||||||| || |||||| | ||||  |||| ||||||||
Db        763 ATGGGCCTCCCTTATCATATTTTTCTCTGTAGTGTTCTTTCACCCTGTTTCACTCTCACT 822

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  ||||||||||||||| ||| |||||||||||| |||
Db        823 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTCACCCAGAGTTTCTATGGAGA 882

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        883 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 942

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        943 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1002

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1003 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1062

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1063 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1122

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||| ||| ||||
Db       1123 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCGGGGTACAT 1182

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1183 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1242

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1243 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1302

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1303 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1362

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1422

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1423 GGACCTCTTGTTTCCAATGTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1482

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | |||||||||| ||||||||||||||||||||||| |||||||||| |
Db       1483 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1542

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1543 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1602

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1662

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db       1663 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1722

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1723 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1782

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1783 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1842

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1902

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              | ||||||||||||||||||||||||||||||  | ||||||||||||||||||||||||
Db       1903 CGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTTA 1962

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| ||||||||||||||| |||||
Db       1963 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGAA 2022

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||| ||||||||||||| ||| |||||||||||||| |||||| | || ||||||||||
Db       2023 ATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2082

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||| ||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       2083 CTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2142

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||| ||||||||||||| |||||||||||||||||
Db       2143 CTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2202

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||| || ||||| ||||      | ||| |  || |
Db       2203 GAAGCTGTGAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2243

RESULT 23
ADE09587
ID   ADE09587 standard; DNA; 3464 BP.
XX
AC   ADE09587;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Novel DNA-related contig nucleotide sequence #309.
XX
KW   novel gene; novel protein; tissue marker; molecular weight marker;
KW   chromosome marker; genetic disorder; contig; ds.
XX
OS   Unidentified.
XX
PN   WO2003054152-A2.
XX
PD   03-JUL-2003.
XX
PF   10-DEC-2002; 2002WO-US039555.
XX
PR   10-DEC-2001; 2001US-0339739P.
PR   11-DEC-2001; 2001US-0339453P.
PR   14-MAR-2002; 2002US-0365091P.
PR   14-MAR-2002; 2002US-0365384P.
PR   12-APR-2002; 2002US-0372381P.
PR   12-APR-2002; 2002US-0372615P.
PR   22-APR-2002; 2002US-00128558.
PR   24-APR-2002; 2002US-0376045P.
XX
PA   (HYSE-) HYSEQ INC.
XX
```

```
PI   Tang YT,  Asundi V,  Goodrich RW,  Ren F,  Zhang J,  Zhao QA,  Wang J;
PI   Ghosh M,  Xue AJ,  Wehrman T,  Weng G,  Zhou P,  Drmanac RT,  Wang Z;
PI   Ma Y,  Wang D,  Chen R,  Xu C,  Boyle BJ;
XX
DR   WPI; 2003-569235/53.
XX
PT   New polynucleotides, useful for expressing recombinant proteins for
PT   analysis, characterization or therapeutic use, or as markers for tissues
PT   in which the corresponding protein is preferentially expressed.
XX
PS   Disclosure; SEQ ID NO 2131; 1177pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of novel
CC   proteins. The DNA and protein sequences of the invention are useful as:
CC   markers for tissues in which the corresponding protein is preferentially
CC   expressed; as molecular weight markers on gels; as chromosome markers or
CC   tags; to identify chromosomes or to map related gene positions; and to
CC   compare with endogenous DNA sequences in patients to identify potential
CC   genetic disorders. The present DNA sequence was used in the
CC   exemplification of the invention.
XX
SQ   Sequence 3464 BP; 1038 A; 913 C; 666 G; 845 T; 0 U; 2 Other;

  Query Match            87.7%;  Score 1298.2;  DB 10;  Length 3464;
  Best Local Similarity  93.0%;  Pred. No. 0;
  Matches 1383;  Conservative    0;  Mismatches   98;  Indels    6;  Gaps    2;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db       1430 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 1489

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       1490 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1549

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db       1550 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1609

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1610 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1669

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1670 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1729

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| |||||||||| |||||||||||||||| ||  |||||||||||||||||||| ||
Db       1730 GGAGTCACTGTCTGTCGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1789

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| ||||||||||||||||||||||||||||||
Db       1790 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGCAATCTCCCAACTGACCCGGGGACAT 1849

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1850 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 1909

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1910 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1969

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| ||||||||||||| ||||||||||||  ||||||||
Db       1970 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 2029

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2030 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 2089

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2090 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 2149

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | |||||||||| ||||||||||||||||||||||| |||||||||| |
Db       2150 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 2209

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       2210 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 2269

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2270 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 2329

Qy        901 ACTGAACAAGATTTATACAATCATGTC----GTACCTAAGCCCCACAACAAAAGAGTACC 956
              ||||||||||||||||||| | |||||          |  |||| |||||||||||||||
Db       2330 ACTGAACAAGATTTATACAGTTATGTCAATATCCTAAAGCCCCCGCAACAAAAGAGTACC 2389

Qy        957 CATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAG 1016
              ||||||||||||||||||  |||||||||| |||||| |||||||||||||||||||| |
Db       2390 CATTCTTCCTTTTGTTATTGGAGCAGGAGTACTAGGCGGACTAGGTACTGGCATTGGCGG 2449

Qy       1017 TATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGA 1076
              ||| ||||||||||||||||||||| ||||||||||||||||| ||||||||||||||||
Db       2450 TATTACAACCTCTACTCAGTTCTACCACAAACTATCTCAAGAACTAAATGGTGACATGGA 2509

Qy       1077 ACAGGT--CACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTA 1134
              ||||||  | |     ||||||||||||||||||||||||||||||||||||||||||||
Db       2510 ACAGGTCGCCCGACTCCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTA 2569

Qy       1135 GTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTA 1194
              ||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db       2570 GTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTA 2629

Qy       1195 TTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTT 1254
               ||||||| |||||| ||||||| ||||||||||||||| |||| |||||||||||||||
Db       2630 CTTTTAGGGGAAGAATGCTGTTACTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTT 2689

Qy       1255 AAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGG 1314
              |||||||||||||||||||||||| |||||||||||||||||| |||||| | || ||||
Db       2690 AAAGAAATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGG 2749

Qy       1315 GGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTA 1374
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||| ||
Db       2750 GGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATA 2809

Qy       1375 ATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCC 1434
              |||||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       2810 ATATTGCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCC 2869

Qy       1435 AGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||| ||||||||||| ||||| ||||      | |||  |  || |
Db       2870 AGAATCGAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2916

RESULT 24
ADF59718
ID   ADF59718 standard; cDNA; 9502 BP.
XX
AC   ADF59718;
XX
DT   12-FEB-2004  (first entry)
XX
DE   Human contig polynucleotide sequence SEQ ID NO:2085.
XX
KW   biological activity; genetic engineering; hybridisation probe; oligomer;
KW   primer; chromosome mapping; gene mapping; recombinant protein production;
KW   human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003080795-A2.
XX
```

```
PD   02-OCT-2003.
XX
PF   09-AUG-2002; 2002WO-US025485.
XX
PR   09-AUG-2001; 2001US-0311261P.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Yang Y,  Wang Z,  Weng G,  Ma Y;
XX
DR   WPI; 2003-876918/81.
DR   P-PSDB; ADF60170.
XX
PT   New polynucleotides, useful as hybridization probes, oligomers or
PT   primers, for chromosome or gene mapping, for the recombinant production
PT   of proteins, and for generating antisense DNA or RNA.
XX
PS   Example 2; SEQ ID NO 2085; 571pp; English.
XX
CC   The present invention describes isolated polynucleotide sequences (I),
CC   which encode polypeptides (II) with biological activity. Also described:
CC   (1) a vector comprising (I); (2) an expression vector comprising (I); (3)
CC   a host cell genetically engineered to comprise (I) which is operatively
CC   associated with a regulatory sequence that modulates expression of (I) in
CC   the host cell; (4) a polypeptide (II) encoded by (I); (5) a composition
CC   comprising the polypeptide of (4) and a carrier; (6) an antibody directed
CC   against the polypeptide of (4); (7) detecting (I) or the polypeptide of
CC   (4) in a sample; (8) identifying a compound that binds to the polypeptide
CC   of (4); (9) producing the polypeptide of (4); and (10) a collection of
CC   polynucleotides comprising at least one of the polynucleotide sequences
CC   (I). The polynucleotides (I) can be used as hybridisation probes,
CC   oligomers or primers, for chromosome or gene mapping, for the recombinant
CC   production of proteins, and for generating antisense DNA or RNA. The
CC   present sequence represents a human contig polynucleotide sequence, which
CC   is used in an example from the present invention.
XX
SQ   Sequence 9502 BP; 2813 A; 2433 C; 1992 G; 2263 T; 0 U; 1 Other;

  Query Match             87.7%;  Score 1298.2;  DB 10;  Length 9502;
  Best Local Similarity   93.0%;  Pred. No. 0;
  Matches 1383;  Conservative    0;  Mismatches   98;  Indels    6;  Gaps    2;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db       1430 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 1489

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       1490 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1549

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db       1550 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1609

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1610 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1669

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1670 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1729

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| |||||||||| |||||||||||||||| ||  |||||||||||||||||||| ||
Db       1730 GGAGTCACTGTCTGTCGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1789

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| ||||||||||||||||||||||||||||||
Db       1790 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGCAATCTCCCAACTGACCCGGGGACAT 1849

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1850 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 1909

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1910 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1969

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1970 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 2029

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2030 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 2089

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2090 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 2149

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       2150 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 2209

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       2210 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 2269

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2270 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 2329

Qy        901 ACTGAACAAGATTTATACAATCATGTC----GTACCTAAGCCCCACAACAAAAGAGTACC 956
              ||||||||||||||||||| | |||||          |  |||| |||||||||||||||
Db       2330 ACTGAACAAGATTTATACAGTTATGTCAATATCCTAAAGCCCCCGCAACAAAAGAGTACC 2389

Qy        957 CATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAG 1016
              ||||||||||||||||||  |||||||||| |||||| |||||||||||||||||||| |
Db       2390 CATTCTTCCTTTTGTTATTGGAGCAGGAGTACTAGGCGGACTAGGTACTGGCATTGGCGG 2449

Qy       1017 TATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGA 1076
              ||| ||||||||||||||||||||| ||||||||||||||||| ||||||||||||||||
Db       2450 TATTACAACCTCTACTCAGTTCTACCACAAACTATCTCAAGAACTAAATGGTGACATGGA 2509

Qy       1077 ACAGGT--CACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTA 1134
              ||||||  | |     ||||||||||||||||||||||||||||||||||||||||||||
Db       2510 ACAGGTCGCCCGACTCCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTA 2569

Qy       1135 GTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTA 1194
              ||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db       2570 GTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTA 2629

Qy       1195 TTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTT 1254
               ||||||| |||||| ||||||| ||||||||||||||| |||| |||||||||||||||
Db       2630 CTTTTAGGGGAAGAATGCTGTTACTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTT 2689

Qy       1255 AAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGG 1314
              |||||||||||||||||||||||| |||||||||||||||||| |||||| | || ||||
Db       2690 AAAGAAATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGG 2749

Qy       1315 GGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTA 1374
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||| ||
Db       2750 GGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATA 2809

Qy       1375 ATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCC 1434
              |||||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       2810 ATATTGCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCC 2869

Qy       1435 AGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||| ||||||||||| ||||| ||||      | |||  |  || |
Db       2870 AGAATCGAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2916


RESULT 25
ABN97978
ID   ABN97978 standard; DNA; 46340 BP.
```

```
XX
AC   ABN97978;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral sequence HI3.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 15; Page 186-199; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 46340 BP; 16104 A; 8738 C; 8434 G; 13064 T; 0 U; 0 Other;

  Query Match             86.6%;  Score 1282;  DB 3;  Length 46340;
  Best Local Similarity   92.3%;  Pred. No. 0;
  Matches 1395;  Conservative    0;  Mismatches   85;  Indels   32;  Gaps    3;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Db      31855 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCCCTTTCACTCTCACT 31914

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      31915 GCACCCCCTCCATGCTGCTGTACAACCAGCAGCTCCCCTTACCAAGAGTTTCTATGAAGA 31974

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | |||||||||  |||||||||||||||||||| ||||||||||| |||| |||||||||
Db      31975 ATGCGGCTTCCCAGAAATATTGATGCCCCATCAAATAGGAGTTTACCTAAAGGAAACTCC 32034

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      32035 ACCTTCACTGCCCACACCCATATGCCCCACAACTGCTATAACTCTGCCACTCTTTGCATG 32094

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      32095 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAAGACTT 32154

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||| |||||||||||||||||||  |||||||||| ||||||||| ||
Db      32155 GGAGCCACTGTCTGTCGGACTTACTTCACCCATACTGGTATGTCTGAGGGGGGTGGAGTT 32214

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| |||||||||| || ||||||||||||||||| ||||
Db      32215 CAAGATCAGGCAAGAGAAAAACATGTAAAGGAAGTAACCTCCCAACTGACCCGGGTACAT 32274

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              |||||||||||||||||||||||||||| ||||| |||||||||||||||||||| ||||
Db      32275 AGCACCCCTAGCCCCTACAAAGGACTAGATCTCTTAAAACTACATGAAACCCTCCATACC 32334

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||| ||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db      32335 CATACTTGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 32394

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              |||||||||||||||||||||||||||||||||||| | |||||||||| ||||||||||
Db      32395 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGTATTTCAGGCCATGCATTTCAATC 32454

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      32455 CCTGTACCTGAACAATGGAACAACTACAGCACAGAAATAAACACCACTTCCGTTTTAGTA 32514

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32515 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 32574

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||| |||||||||||| |||||||||||||||||||||||| ||||||||||||
Db      32575 AGCAATACTGTAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACGA 32634

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||
Db      32635 ATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGTGGTACCTTAGCCTATCGTTGTTTG 32694

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||| |||||||| |||
Db      32695 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCC-ATGACCATTTAC 32753

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||| |||| |||||||||||||||||||||||||||| ||||
Db      32754 ACTGAACAAGATTTATACAATTATGTTGTACCTAAGCCCCACAACAAAAGAGTACTCATT 32813

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||| ||||||||||||||||  |||||||| ||||||||||| ||| |
Db      32814 CTTCCTTTTGTTATCGGAGCAGGAGTGCTAGGTGGACTAGGTTCTGGCATTGGCGGTACC 32873

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| | |||||||||||||||  |
Db      32874 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTCAATGGTGACATGGAATGG 32933

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||  | |||||||||||||||||||||||||||||||||| ||||||| |||||||||||
Db      32934 GTTGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTTCCTAGCATCAGTAGTCCTT 32993

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGAACCTGTTTATTTTTA 1200
              |||||| ||||||||||||||||||||||| |  ||||||||||| ||||||||||||||
Db      32994 CAAAATTGAAGAGCTTTAGACTTGCTAACCTCTGAAAGAGGGGAAGCTGTTTATTTTTA 33053

Qy       1201 GGAGAAGAACGCTG------------------------------TTATTATGTTAATCAA 1230
              || |||||| | ||                              ||||||||||||||||
Db      33054 GGGGAAGAATGTTGTTATTATGTTATTTAGCGGAAGAATGTTGTTATTATGTTAATCAA 33113

Qy       1231 TCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAG 1290
              ||| |||||||||| ||||||||| |||||||||||||| |||||||| ||||| |||||
Db      33114 TCCTGAATTGTCACAGAGAAAGTTGAAGAAATTCGAGATTGAATACAACGTAGAACAGAG 33173

Qy       1291 GAGCTTC-AAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCC 1349
              ||||||| ||||||||  || ||||||||||||||||||||||||||||||| |||||||
Db      33174 GAGCTTCAAAAACACCAGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCC 33233

Qy       1350 CTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAA 1409
              |||||||||| |||||||||||||||||||| ||||||||||||||||||||||||||||
Db        33234 CTTCTTAGGATCTCTAGCAGCTCTAATATTGATACTCCTCTTTGGACCCTGTATCTTTAA 33293

Qy         1410 CCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACA 1469
                ||||||||||||||||||||||||||||||  ||| |||||||||||| || ||  | ||
Db        33294 CCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGTTGTAAAGCTACAAATCGTTCTTCA 33353

Qy         1470 AATGGAACCCCA 1481
                ||||||||||||
Db        33354 AATGGAACCCCA 33365


RESULT 26
ABN97931/c
ID   ABN97931 standard; DNA; 1799 BP.
XX
AC   ABN97931;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral sequence HE3.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 3; Page 134; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 1799 BP; 488 A; 331 C; 482 G; 498 T; 0 U; 0 Other;

  Query Match             86.3%;  Score 1278.8;  DB 3;  Length 1799;
  Best Local Similarity   92.1%;  Pred. No. 0;
  Matches 1393;  Conservative    0;  Mismatches   87;  Indels   32;  Gaps    3;

Qy            1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
                ||||||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Db         1762 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCCCTTTCACTCTCACT 1703

Qy           61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
                ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         1702 GCACCCCCTCCATGCTGCTGTACAACCAGCAGCTCCCCTTACCAAGAGTTTCTATGAAGA 1643

Qy          121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
                | |||||||||  |||||||||||||||||||| ||||||||||| |||| |||||||||
Db         1642 ATGCGGCTTCCCAGAAATATTGATGCCCCATCAAATAGGAGTTTACCTAAAGGAAACTCC 1583

Qy          181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
                |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         1582 ACCTTCACTGCCCACACCCATATGCCCCACAACTGCTATAACTCTGCCACTCTTTGCATG 1523

Qy          241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
                |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db         1522 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAAGACTT 1463

Qy          301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
                ||||||||||||||| |||||||||||||||||||  |||||||||| ||||||||| ||
Db         1462 GGAGCCACTGTCTGTCGGACTTACTTCACCCATACTGGTATGTCTGAGGGGGGTGGAGTT 1403

Qy          361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
                |||| |||||||||||||||||| |||||||||| || ||||||||||||||||| ||||
Db         1402 CAAGATCAGGCAAGAGAAAAACATGTAAAGGAAGTAACCTCCCAACTGACCCGGGTACAT 1343

Qy          421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                |||||||||||||||||||||||||||| ||||| |||||||||||||||||||| ||||
Db         1342 AGCACCCCTAGCCCCTACAAAGGACTAGATCTCTTAAAACTACATGAAACCCTCCATACC 1283

Qy          481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
                |||||| ||||||| |||||||||||||||||||||||| |||||||||||||||| | |
Db         1282 CATACTTGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGTC 1223

Qy          541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
                ||||||||||||||||||||| |||||||||||||| | |||||||||| ||||||||||
Db         1222 CAAAACCCTACTAACTGTTGGTTGTGCCTCCCCCTGTATTTCAGGCCATGCATTTCAATC 1163

Qy          601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
                ||||| ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         1162 CCTGTACCTGAACAATGGAACAACTACAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1103

Qy          661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1102 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1043

Qy          721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
                ||||||||| |||||||||||| |||||||||||||||||||||||| ||||||||||||
Db         1042 AGCAATACTGTAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACGA 983

Qy          781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
                ||||||||||||||||||||||||||||||||||||||||||| |||||||| |||||||
Db          982 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTTAGCCTATCGTTGTTTG 923

Qy          841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
                ||||||||||||||||||||||||||||||||||||||||||||||| |||||||| |||
Db          922 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCC-ATGACCATTTAC 864

Qy          901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
                |||||||||||||||||||| |||| ||||||||||||||||||||||||||||| ||||
Db          863 ACTGAACAAGATTTATACAATTATGTTGTACCTAAGCCCCACAACAAAAGAGTACTCATT 804

Qy          961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
                ||||||||||||||| ||||||||||||||||  |||||||| ||||||||||| ||| |
Db          803 CTTCCTTTTGTTATCGGAGCAGGAGTGCTAGGTGGACTAGGTTCTGGCATTGGCGGTACC 744

Qy         1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
                ||||||||||||||||||||||||||||||||||||||| | |||||||||||||||  |
Db          743 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTCAATGGTGACATGGAATGG 684

Qy         1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
                ||  | |||||||||||||||||||||||||||||||||| ||||||| |||||||||||
Db          683 GTTGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTTCCTAGCATCAGTAGTCCTT 624

Qy         1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
                |||||| ||||||||||||||||||||||| |  |||||||||||| |||||||||||||
Db          623 CAAAATTGAAGAGCTTTAGACTTGCTAACCTCTGAAAGAGGGGGAAGCTGTTTATTTTTA 564
```

```
Qy       1201 GGAGAAGAACGCTG------------------------------TTATTATGTTAATCAA 1230
              || |||||| | ||                              ||||||||||||||||
Db        563 GGGGAAGAATGTTGTTATTATGTTATTTTAGCGGAAGAATGTTGTTATTATGTTAATCAA 504

Qy       1231 TCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAG 1290
              ||| |||||||||| ||||||||| |||||||||||||| |||||||| ||||| |||||
Db        503 TCCTGAATTGTCACAGAGAAAGTTGAAGAAATTCGAGATTGAATACAACGTAGAACAGAG 444

Qy       1291 GAGCTTC-AAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCC 1349
              ||||||| ||||||||  || ||||||||||||||||||||||||||||||| |||||||
Db        443 GAGCTTCAAAAACACCAGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCC 384

Qy       1350 CTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAA 1409
              |||||||||| |||||||||||||||||||| ||||||||||||||||||||||||||||
Db        383 CTTCTTAGGATCTCTAGCAGCTCTAATATTGATACTCCTCTTTGGACCCTGTATCTTTAA 324

Qy       1410 CCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACA 1469
              ||||||||||||||||||||||||||||||  ||| |||||||||||| || ||  | ||
Db        323 CCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGTTGTAAAGCTACAAATCGTTCTTCA 264

Qy       1470 AATGGAACCCCA 1481
              ||||||||||||
Db        263 AATGGAACCCCA 252


RESULT 27
ABN97930
ID   ABN97930 standard; DNA; 2784 BP.
XX
AC   ABN97930;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral sequence HE2.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 3; Page 133; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 2784 BP; 816 A; 755 C; 535 G; 678 T; 0 U; 0 Other;

  Query Match             83.7%;  Score 1239;  DB 3;  Length 2784;
  Best Local Similarity   91.0%;  Pred. No. 0;
  Matches 1364;  Conservative    0;  Mismatches  100;  Indels   35;  Gaps    3;

Qy          2 TGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACTG 61
              |||||||||||||||||| ||||||||||||||||||||||||||||||| |||||||||
Db        903 TGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCCCTTTCACTCTCACTG 962

Qy         62 CACCCCCTCCATGCTGCTGTAC-------------------AACCAGTAGCTCCCCTTAC 102
              |||||| |||||||  ||| ||                     ||||||||||||||||
Db        963 CACCCCGTCCATGCCACTGCACCCCGTCCATGCCCCGTCCATGCCAGTAGCTCCCCTTAG 1022

Qy        103 CAAGAGTTTCTATGAAGAACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGT 162
              |||||||||||||| |||| || || ||| |||||||||||||||||||  |||||||||
Db       1023 CAAGAGTTTCTATGGAGAATGCAGCGTCCCGGAAATATTGATGCCCCATTGTATAGGAGT 1082

Qy        163 TTATCTAAGGGAAACTCCACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAAC 222
              ||||||||||||| | |||||||||||||||||||||||||||||| |||||||||||||
Db       1083 TTATCTAAGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCCACAACTGCTATAAC 1142

Qy        223 TCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCT 282
              |||||||||||||||||||||||||||||||||||||||||||| |||| |||||||||
Db       1143 TCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAACGATTAATCCC 1202

Qy        283 AGTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATG 342
              |||||||||||||||||||||             ||||| ||||||| ||||||||||||
Db       1203 AGTTGTCCTGGAGGACTTGGA-------------GGACTCACTTCACTCATACCAGTATG 1249

Qy        343 TCTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCC 402
              ||||||||||||||| |||||| |||||||| |||||||||  ||||||||| |||||||
Db       1250 TCTGATGGGGGTGGAGTTCAAGATCAGGCAACAGAAAAACACATAAAGGAAGTAATCTCC 1309

Qy        403 CAACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTA 462
              ||||||||| ||| ||||||||||||| |||||||||||||||||| |||||||||||||
Db       1310 CAACTGACCTGGGTACATAGCACCCCTGGCCCCTACAAAGGACTAGATCTCTCAAAACTA 1369

Qy        463 CATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGG 522
              ||||||||||||| |||||||||| ||||||| |||||||||||||||||||| ||| ||
Db       1370 CATGAAACCCTCCATACCCATACTGGCCTGGTAAGCCTATTTAATACCACCCTGACTGGG 1429

Qy        523 CTCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTC 582
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1430 CTCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTT 1489

Qy        583 AGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAAC 642
              ||||||||||||||||||||| | ||||||||||||||||||||||||||||||||||||
Db       1490 AGGCCATACATTTCAATCCCTATACCTGAACAATGGAACAACTTCAGCACAGAAATAAAC 1549

Qy        643 ACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAAC 702
              |||||||| ||||||||||| ||||   ||||||||||||||||||||||||||||||||
Db       1550 ACCACTTCTGTTTTAGTAGGTCCTC---TTTCCAATCTGGAAATAACCCATACCTCAAAC 1606

Qy        703 CTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGG 762
              |||||||||||||||||||||||||||||||||||| ||| |||||||||||||||||||
Db       1607 CTCACCTGTGTAAAATTTAGCAATACTATAGACACAGCCAACTCCCAATGCATCAGGTGG 1666

Qy        763 GTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACC 822
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 GTAACTCCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACC 1726

Qy        823 TCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTG 882
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db       1727 TCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTGTGTGCTTCCTCTCATTCTTAGTG 1786

Qy        883 CCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCAC 942
               |||||||| ||||||||||||||||||||||||||||||||||| |||||||||||| |
Db       1787 GCCCCTATGCCCATCTACACTGAACAAGATTTATACAATCATGTCATACCTAAGCCCCGC 1846

Qy        943 AACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGT 1002
              ||||||||||||||||||||||||||||||||  ||||||||||||||||| || ||| |
Db       1847 AACAAAAGAGTACCCATTCTTCCTTTTGTTATTGGAGCAGGAGTGCTAGGCGGAGTAGCT 1906

Qy       1003 ACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATA 1062
              |||||||||||| |||||||||||||||||||||||||||||||||| ||||||||| ||
Db       1907 ACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTGTCTCAAGAACTA 1966

Qy       1063 AATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCC 1122
              |||||||||||||||  |||| ||||  ||||||||||||||||||||||||||||||||
Db       1967 AATGGTGACATGGAATGGGTCGCTGATACCCTGGTCACCTTGCAAGATCAACTTAACTCC 2026

Qy       1123 CTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGG 1182
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |||| |||
Db       2027 CTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGG 2086

Qy       1183 GGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTC 1242
              ||||||| |||||||||||  ||| || |||||| |||||||||||||||| ||||  ||
Db       2087 GGAACCTTTTTATTTTTAGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATC 2146

Qy       1243 ACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAAC 1302
              || ||||||||||||||||||| || ||||||| || |||||||| ||||||| ||||||
Db       2147 ACCGAGAAAGTTAAAGAAATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAAC 2206

Qy       1303 ACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCT 1362
              || | || ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       2207 ACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCT 2266

Qy       1363 CTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAG 1422
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 CTAGCAGCTATAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAG 2326

Qy       1423 TTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||| ||||||||| ||||| ||||| ||||| || ||  | |||||||| |||||
Db       2327 TTTGTCTTTTCCAGAATCGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCA 2385


RESULT 28
AAI14608
ID   AAI14608 standard; DNA; 1894 BP.
XX
AC   AAI14608;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #4541 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 4541; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

  Query Match             76.4%;  Score 1131.6;  DB 4;  Length 1894;
  Best Local Similarity   92.4%;  Pred. No. 0;
  Matches 1218;  Conservative    0;  Mismatches   84;  Indels   16;  Gaps    2;

Qy        164 TATCTAAGGGAAACTCCACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACT 223
              |||||||||||| | |||||||||||||||||||||||||||||| ||||||||||||||
Db          1 TATCTAAGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCCACAACTGCTATAACT 60

Qy        224 CTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTA 283
              ||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||| |
Db         61 CTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAACGATTAATCCCA 120

Qy        284 GTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGT 343
              ||||||||||||||||||||             ||||| ||||||| |||||||||||||
Db        121 GTTGTCCTGGAGGACTTGGA-------------GGACTCACTTCACTCATACCAGTATGT 167

Qy        344 CTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCC 403
              |||||||||||||| |||||| |||||||| |||||||||  ||||||||| ||||||||
Db        168 CTGATGGGGGTGGAGTTCAAGATCAGGCAACAGAAAAACACATAAAGGAAGTAATCTCCC 227

Qy        404 AACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTAC 463
              |||||||| ||| ||||||||||||| |||||||||||||||||| ||||||||||||||
Db        228 AACTGACCTGGGTACATAGCACCCCTGGCCCCTACAAAGGACTAGATCTCTCAAAACTAC 287

Qy        464 ATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGC 523
              |||||||||||| |||||||||| ||||||| |||||||||||||||||||| ||| |||
Db        288 ATGAAACCCTCCATACCCATACTGGCCTGGTAAGCCTATTTAATACCACCCTGACTGGGC 347

Qy        524 TCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCA 583
              ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||| |
Db        348 TCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTTA 407

Qy        584 GGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACA 643
              |||||||||||||||||||| | |||||||||||||||||||||||||||||||||||||
Db        408 GGCCATACATTTCAATCCCTATACCTGAACAATGGAACAACTTCAGCACAGAAATAAACA 467

Qy        644 CCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACC 703
              ||||||| ||||||||||| ||||   |||||||||||||||||||||||||||||||||
Db        468 CCACTTCTGTTTTAGTAGGTCCTC---TTTCCAATCTGGAAATAACCCATACCTCAAACC 524

Qy        704 TCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGG 763
              |||||||||||||||||||||||||||||||||| ||| |||||||||||||||||||||
Db        525 TCACCTGTGTAAAATTTAGCAATACTATAGACACAGCCAACTCCCAATGCATCAGGTGGG 584

Qy        764 TAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCT 823
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 TAACTCCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCT 644

Qy        824 CAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGC 883
              ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        645 CAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTGTGTGCTTCCTCTCATTCTTAGTGG 704

Qy        884 CCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACA 943
              |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||| ||
Db        705 CCCCTATGCCCATCTACACTGAACAAGATTTATACAATCATGTCATACCTAAGCCCCGCA 764

Qy        944 ACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTA 1003
              |||||||||||||||||||||||||||||||  ||||||||||||||||| || ||| ||
Db          765 ACAAAAGAGTACCCATTCTTCCTTTTGTTATTGGAGCAGGAGTGCTAGGCGGAGTAGCTA 824

Qy         1004 CTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAA 1063
                ||||||||||| |||||||||||||||||||||||||||||||||| ||||||||| |||
Db          825 CTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTGTCTCAAGAACTAA 884

Qy         1064 ATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCC 1123
                ||||||||||||||  |||| ||||  |||||||||||||||||||||||||||||||||
Db          885 ATGGTGACATGGAATGGGTCGCTGATACCCTGGTCACCTTGCAAGATCAACTTAACTCCC 944

Qy         1124 TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGG 1183
                |||||||||||||||||||||||||||||||||||||||||||||||||  |||| ||||
Db          945 TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGG 1004

Qy         1184 GAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCA 1243
                |||||| |||||||||||  ||| || |||||| |||||||||||||||| ||||  |||
Db         1005 GAACCTTTTTATTTTTAGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCA 1064

Qy         1244 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACA 1303
                | ||||||||||||||||||| || ||||||| || |||||||| ||||||| |||||||
Db         1065 CCGAGAAAGTTAAAGAAATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACA 1124

Qy         1304 CCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTC 1363
                | | || ||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db         1125 CTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTC 1184

Qy         1364 TAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGT 1423
                |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1185 TAGCAGCTATAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGT 1244

Qy         1424 TTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
                |||||| ||||||||| ||||| ||||| ||||| || ||  | |||||||| |||||
Db         1245 TTGTCTTTTCCAGAATCGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCA 1302

RESULT 29
ABA56337
ID   ABA56337 standard; DNA; 1894 BP.
XX
AC   ABA56337;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #4642.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 1; SEQ ID NO 4642; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

  Query Match             76.4%;  Score 1131.6;  DB 4;  Length 1894;
  Best Local Similarity   92.4%;  Pred. No. 0;
  Matches 1218;  Conservative    0;  Mismatches   84;  Indels   16;  Gaps    2;

Qy          164 TATCTAAGGGAAACTCCACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACT 223
                |||||||||||| | |||||||||||||||||||||||||||||| ||||||||||||||
Db            1 TATCTAAGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCCACAACTGCTATAACT 60

Qy          224 CTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTA 283
                ||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||| |
Db           61 CTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAACGATTAATCCCA 120

Qy          284 GTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGT 343
                ||||||||||||||||||||             ||||| ||||||| |||||||||||||
Db          121 GTTGTCCTGGAGGACTTGGA-------------GGACTCACTTCACTCATACCAGTATGT 167

Qy          344 CTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCC 403
                |||||||||||||| |||||| |||||||| |||||||||  ||||||||| ||||||||
Db          168 CTGATGGGGGTGGAGTTCAAGATCAGGCAACAGAAAAACACATAAAGGAAGTAATCTCCC 227

Qy          404 AACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTAC 463
                |||||||| ||| ||||||||||||| |||||||||||||||||| ||||||||||||||
Db          228 AACTGACCTGGGTACATAGCACCCCTGGCCCCTACAAAGGACTAGATCTCTCAAAACTAC 287

Qy          464 ATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGC 523
                |||||||||||| ||||||||||| ||||||| ||||||||||||||||||| ||| |||
Db          288 ATGAAACCCTCCATACCCATACTGGCCTGGTAAGCCTATTTAATACCACCCTGACTGGGC 347

Qy          524 TCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCA 583
                ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||| |
Db          348 TCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTTA 407

Qy          584 GGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACA 643
                |||||||||||||||||||| | |||||||||||||||||||||||||||||||||||||
Db          408 GGCCATACATTTCAATCCCTATACCTGAACAATGGAACAACTTCAGCACAGAAATAAACA 467

Qy          644 CCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACC 703
                ||||||| ||||||||||| ||||   |||||||||||||||||||||||||||||||||
Db          468 CCACTTCTGTTTTAGTAGGTCCTC---TTTCCAATCTGGAAATAACCCATACCTCAAACC 524

Qy          704 TCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGG 763
                ||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db          525 TCACCTGTGTAAAATTTAGCAATACTATAGACACAGCCAACTCCCAATGCATCAGGTGGG 584

Qy          764 TAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCT 823
                |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          585 TAACTCCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCT 644

Qy          824 CAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGC 883
                ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db          645 CAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTGTGTGCTTCCTCTCATTCTTAGTGG 704

Qy          884 CCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACA 943
                |||||||| ||||||||||||||||||||||||||||||||||| |||||||||||| ||
Db          705 CCCCTATGCCCATCTACACTGAACAAGATTTATACAATCATGTCATACCTAAGCCCCGCA 764

Qy          944 ACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTA 1003
                |||||||||||||||||||||||||||||||  ||||||||||||||||| || ||| ||
Db          765 ACAAAAGAGTACCCATTCTTCCTTTTGTTATTGGAGCAGGAGTGCTAGGCGGAGTAGCTA 824

Qy         1004 CTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAA 1063
```

```
              ||||||||||| |||||||||||||||||||||||||||||||||| ||||||||| |||
Db        825 CTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTGTCTCAAGAACTAA 884

Qy       1064 ATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCC 1123
              ||||||||||||||  |||| ||||  |||||||||||||||||||||||||||||||||
Db        885 ATGGTGACATGGAATGGGTCGCTGATACCCTGGTCACCTTGCAAGATCAACTTAACTCCC 944

Qy       1124 TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGG 1183
              |||||||||||||||||||||||||||||||||||||||||||||||||  |||| ||||
Db        945 TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGG 1004

Qy       1184 GAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCA 1243
              |||||| |||||||||||  ||| || |||||| |||||||||||||||| ||||  |||
Db       1005 GAACCTTTTTATTTTTAGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCA 1064

Qy       1244 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACA 1303
              | ||||||||||||||||||| || ||||||| || |||||||| ||||||| |||||||
Db       1065 CCGAGAAAGTTAAAGAAATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACA 1124

Qy       1304 CCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTC 1363
              | | || ||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db       1125 CTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTC 1184

Qy       1364 TAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGT 1423
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 TAGCAGCTATAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGT 1244

Qy       1424 TTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||| ||||||||| ||||| ||||| ||||| || ||  | |||||||| |||||
Db       1245 TTGTCTTTTCCAGAATCGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCA 1302


RESULT 30
AAI35980
ID   AAI35980 standard; DNA; 1894 BP.
XX
AC   AAI35980;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #4666 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000663.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human placenta.
XX
PS   Claim 25; SEQ ID NO 4666; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
```

```
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

  Query Match             76.4%;  Score 1131.6;  DB 4;  Length 1894;
  Best Local Similarity   92.4%;  Pred. No. 0;
  Matches 1218;  Conservative    0;  Mismatches   84;  Indels   16;  Gaps    2;

Qy        164 TATCTAAGGGAAACTCCACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACT 223
              |||||||||||| | ||||||||||||||||||||||||||||||| |||||||||||||
Db          1 TATCTAAGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCCACAACTGCTATAACT 60

Qy        224 CTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTA 283
              ||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||| |
Db         61 CTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAACGATTAATCCCA 120

Qy        284 GTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGT 343
              ||||||||||||||||||||             ||||| ||||||| |||||||||||||
Db        121 GTTGTCCTGGAGGACTTGGA-------------GGACTCACTTCACTCATACCAGTATGT 167

Qy        344 CTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCC 403
              |||||||||||||| |||||| |||||||| |||||||||  ||||||||| ||||||||
Db        168 CTGATGGGGGTGGAGTTCAAGATCAGGCAACAGAAAAACACATAAAGGAAGTAATCTCCC 227

Qy        404 AACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTAC 463
              |||||||| ||| ||||||||||||| ||||||||||||||||||| ||||||||||||||
Db        228 AACTGACCTGGGTACATAGCACCCCTGGCCCCTACAAAGGACTAGATCTCTCAAAACTAC 287

Qy        464 ATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGC 523
              |||||||||||| |||||||||| ||||||| |||||||||||||||||||| ||| |||
Db        288 ATGAAACCCTCCATACCCATACTGGCCTGGTAAGCCTATTTAATACCACCCTGACTGGGC 347

Qy        524 TCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCA 583
              ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||| |
Db        348 TCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTTA 407

Qy        584 GGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACA 643
              |||||||||||||||||||| | |||||||||||||||||||||||||||||||||||||
Db        408 GGCCATACATTTCAATCCCTATACCTGAACAATGGAACAACTTCAGCACAGAAATAAACA 467

Qy        644 CCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACC 703
              ||||||| ||||||||||| ||||   |||||||||||||||||||||||||||||||||
Db        468 CCACTTCTGTTTTAGTAGGTCCTC---TTTCCAATCTGGAAATAACCCATACCTCAAACC 524

Qy        704 TCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGG 763
              ||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db        525 TCACCTGTGTAAAATTTAGCAATACTATAGACACAGCCAACTCCCAATGCATCAGGTGGG 584

Qy        764 TAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCT 823
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 TAACTCCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCT 644

Qy        824 CAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGC 883
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        645 CAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTGTGTGCTTCCTCTCATTCTTAGTGG 704

Qy        884 CCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACA 943
              |||||||| ||||||||||||||||||||||||||||||||||||| |||||||||| ||
Db        705 CCCCTATGCCCATCTACACTGAACAAGATTTATACAATCATGTCATACCTAAGCCCCGCA 764

Qy        944 ACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTA 1003
              |||||||||||||||||||||||||||||||  ||||||||||||||||| || ||| ||
Db        765 ACAAAAGAGTACCCATTCTTCCTTTTGTTATTGGAGCAGGAGTGCTAGGCGGAGTAGCTA 824

Qy       1004 CTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAA 1063
              ||||||||||| |||||||||||||||||||||||||||||||||| ||||||||| |||
Db        825 CTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTGTCTCAAGAACTAA 884

Qy       1064 ATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCC 1123
              ||||||||||||||  |||| ||||  |||||||||||||||||||||||||||||||||
Db         885 ATGGTGACATGGAATGGGTCGCTGATACCCTGGTCACCTTGCAAGATCAACTTAACTCCC 944

Qy        1124 TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGG 1183
               |||||||||||||||||||||||||||||||||||||||||||||||||  |||| ||||
Db         945 TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGG 1004

Qy        1184 GAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCA 1243
               |||||| |||||||||||  ||| || |||||| |||||||||||||||| ||||  |||
Db        1005 GAACCTTTTTATTTTTAGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCA 1064

Qy        1244 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACA 1303
               | ||||||||||||||||||| || ||||||| || |||||||| ||||||| |||||||
Db        1065 CCGAGAAAGTTAAAGAAATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACA 1124

Qy        1304 CCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTC 1363
               | | || ||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        1125 CTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTC 1184

Qy        1364 TAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGT 1423
               |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1185 TAGCAGCTATAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGT 1244

Qy        1424 TTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               |||||| ||||||||| ||||| ||||| ||||| || ||  | |||||||| |||||
Db        1245 TTGTCTTTTCCAGAATCGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCA 1302

Search completed: January 22, 2005, 16:40:33
Job time : 695.669 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 12:03:14 ; Search time 137.138 Seconds
                                             (without alignments)
                                             7676.074 Million cell updates/sec

Title:          US-09-319-156B-9
Perfect score:  1481
Sequence:       1 atggccctcccttatcatac..........gtcttacaaatggaacccca 1481

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters:      1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                 2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                 3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                 4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                 5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                 6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                                    SUMMARIES
                    %
Result              Query
   No.     Score    Match Length DB  ID                         Description
  -------------------------------------------------------------------------------
     1      1481    100.0   1481  4  US-08-979-847B-105          Sequence 105, App
     2    1335.4     90.2   2946  3  US-09-175-928-3             Sequence 3, Appli
     3       302     20.4   1329  4  US-08-979-847B-108          Sequence 108, App
     4     173.8     11.7    542  1  US-08-686-878A-48           Sequence 48, Appl
     5     173.8     11.7    542  1  US-08-721-489-2             Sequence 2, Appli
     6     145.6      9.8   1722  3  US-08-691-563C-58           Sequence 58, Appl
     7     145.6      9.8   1722  4  US-09-374-766-58            Sequence 58, Appl
     8     145.6      9.8   1722  4  US-08-979-847B-54           Sequence 54, Appl
     9     144.6      9.8   2908  4  US-09-799-451-249           Sequence 249, App
    10     131.2      8.9   1859  3  US-08-691-563C-46           Sequence 46, Appl
    11     131.2      8.9   1859  4  US-09-374-766-46            Sequence 46, Appl
    12     131.2      8.9   1859  4  US-08-979-847B-42           Sequence 42, Appl
    13     130.2      8.8   1001  4  US-09-671-317-406           Sequence 406, App
    14     116.8      7.9   1704  1  US-08-007-282B-1            Sequence 1, Appli
    15     114.8      7.8  80246  3  US-09-078-294-4             Sequence 4, Appli
    16     114.2      7.7   2518  3  US-09-011-745-1             Sequence 1, Appli
    17     114.2      7.7   5865  3  US-09-011-745-8             Sequence 8, Appli
    18     112.2      7.6   4776  4  US-09-555-352-1             Sequence 1, Appli
    19     106.4      7.2   2809  4  US-09-620-312D-171          Sequence 171, App
    20     104.8      7.1   2342  4  US-09-904-615-12            Sequence 12, Appl
    21      90.8      6.1  80595  3  US-09-078-294-3             Sequence 3, Appli
    22        89      6.0    464  3  US-09-120-653D-6            Sequence 6, Appli
    23        89      6.0   3910  3  US-09-120-653D-1            Sequence 1, Appli
    24      70.2      4.7   2002  4  US-09-315-127-7             Sequence 7, Appli
    25        70      4.7    149  4  US-09-513-999C-24228        Sequence 24228, A
    26        70      4.7   3925  3  US-09-011-745-9             Sequence 9, Appli
    27      68.4      4.6    635  4  US-08-979-847B-102          Sequence 102, App
    28      68.4      4.6   6028  3  US-09-011-745-5             Sequence 5, Appli
    29      68.4      4.6   8060  3  US-08-766-528-1             Sequence 1, Appli
    30      68.4      4.6   8060  4  US-09-661-858-1             Sequence 1, Appli
    31        68      4.6  10970  3  US-08-716-351A-5            Sequence 5, Appli
    32        67      4.5   1911  1  US-08-258-420-8             Sequence 8, Appli
    33        67      4.5   1965  1  US-08-258-420-9             Sequence 9, Appli
    34        67      4.5   2001  3  US-08-850-961-13            Sequence 13, Appl
    35        67      4.5   2001  3  US-09-479-776-13            Sequence 13, Appl
    36        67      4.5   2001  4  US-09-315-127-10            Sequence 10, Appl
c   37        67      4.5   8202  1  US-08-258-420-13            Sequence 13, Appl
    38      66.8      4.5   6312  3  US-09-011-745-7             Sequence 7, Appli
    39      66.8      4.5   8132  3  US-08-766-528-3             Sequence 3, Appli
    40      66.8      4.5   8132  4  US-09-661-858-3             Sequence 3, Appli
c   41      65.6      4.4    371  4  US-09-621-976-14048         Sequence 14048, A
    42      65.4      4.4   6061  3  US-09-011-745-6             Sequence 6, Appli
    43      64.2      4.3   1989  3  US-09-232-278A-5            Sequence 5, Appli
    44      64.2      4.3   2499  1  US-08-105-483-310           Sequence 310, App
    45      64.2      4.3   2499  1  US-08-709-209-310           Sequence 310, App

                                    ALIGNMENTS

RESULT 1
US-08-979-847B-105
; Sequence 105, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,847B
;            FILING DATE: 26-No. 6582703-1997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;  INFORMATION FOR SEQ ID NO: 105:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1481 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 105:
US-08-979-847B-105
```

```
  Query Match             100.0%;  Score 1481;  DB 4;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 2
US-09-175-928-3
; Sequence 3, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT: McCoy, John M.
;  APPLICANT: LaVallie, Edward R.
;  APPLICANT: Collins-Racie, Lisa A.
;  APPLICANT: Evans, Cheryl
;  APPLICANT: Merberg, David
;  APPLICANT: Treacy, Maurice
;  APPLICANT: Mi, Sha
;  APPLICANT: Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6006B.AJ172A
;  CURRENT APPLICATION NUMBER: US/09/175,928A
;  CURRENT FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-175-928-3

  Query Match             90.2%;  Score 1335.4;  DB 3;  Length 2946;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        928 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 987

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        988 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1047

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db     1048 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1107

Qy      181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
            ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     1108 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1167

Qy      241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
            |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     1168 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1227

Qy      301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
            |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db     1228 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1287

Qy      361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
            |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db     1288 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1347

Qy      421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
             ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     1348 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1407

Qy      481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
            |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db     1408 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1467

Qy      541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
            ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db     1468 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1527

Qy      601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
            ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1528 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1587

Qy      661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1588 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1647

Qy      721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
            |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db     1648 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1707

Qy      781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
            |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     1708 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1767

Qy      841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1768 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1827

Qy      901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
            ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db     1828 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1887

Qy      961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
            ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db     1888 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1947

Qy     1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
            ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db     1948 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2007

Qy     1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
            ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2008 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2067

Qy     1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
            ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db     2068 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2127

Qy     1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
            || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db     2128 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2187

Qy     1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
            |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db     2188 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2247

Qy     1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
            ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db     2248 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2307

Qy     1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db     2308 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2367

Qy     1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
            ||||||||||| ||||| ||||      | ||| |  || |
Db     2368 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2408

RESULT 3
US-08-979-847B-108
; Sequence 108, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,847B
;            FILING DATE: 26-No. 6582703-1997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;  INFORMATION FOR SEQ ID NO: 108:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1329 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 108:
US-08-979-847B-108

  Query Match            20.4%;  Score 302;  DB 4;  Length 1329;
  Best Local Similarity  92.7%;  Pred. No. 1.1e-87;
  Matches  317;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;
```

```
Qy        1140 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 1199
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy        1200 AGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGA 1259
               ||| |||||| ||||||| ||||||||||||||  ||||  | | |||||||||||||||||
Db          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy        1260 AATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCT 1319
               |||| ||||||||||| |||||||||||||||| ||||||||||| | || |||||||||
Db         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy        1320 CCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATT 1379
               ||||||||||||||||||||||  |||||||||||||||||||||||||||| |||||||
Db         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Qy        1380 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAAT 1439
                |||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||
Db         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy        1440 TGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               |||||||||||||||||| || ||  | ||||||||||||||
Db         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCA 342
```

```
RESULT 4
US-08-686-878A-48
; Sequence 48, Application US/08686878A
; Patent No. 5708157
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Evans, Cheryl
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/686,878A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  542 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-686-878A-48

  Query Match             11.7%;  Score 173.8;  DB 1;  Length 542;
  Best Local Similarity   91.5%;  Pred. No. 3.5e-46;
  Matches  184;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db         342 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 401

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         402 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 461

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db         462 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 521

Qy         181 ACCTTCACTGCCCACACCCAT 201
               |||||||||||||||||||||
Db         522 ACCTTCACTGCCCACACCCAT 542
```

```
RESULT 5
US-08-721-489-2
; Sequence 2, Application US/08721489
; Patent No. 5786465
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/721,489
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  542 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-721-489-2

  Query Match             11.7%;  Score 173.8;  DB 1;  Length 542;
  Best Local Similarity   91.5%;  Pred. No. 3.5e-46;
  Matches  184;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db         342 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 401
```

```
Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         402 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 461

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db         462 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 521

Qy         181 ACCTTCACTGCCCACACCCAT 201
               |||||||||||||||||||||
Db         522 ACCTTCACTGCCCACACCCAT 542


RESULT 6
US-08-691-563C-58
; Sequence 58, Application US/08691563C
; Patent No. 6001987
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563C
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 base pairs
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-691-563C-58

  Query Match            9.8%;  Score 145.6;  DB 3;  Length 1722;
  Best Local Similarity   55.7%;  Pred. No. 1.1e-36;
  Matches  302;  Conservative    0;  Mismatches  234;  Indels    6;  Gaps    1;

Qy         940 CACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTA 999
               ||   || |    |||   |  |  |  |   ||||  |   |||| ||           |
Db         346 CAGTGCAGAGCCATACAACTAATATCCCTATTTATAGGGTTAGGAATGGCTACTGCTACA 405

Qy        1000 GGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAA 1059
               || ||||| || | | || |  | || ||       | |||| | | ||| ||  |  |
Db         406 GGAACTGGAATAGCCGGTTTATCTACTTCATTATCCTACTACCATACACTCTCAAAGAAT 465
```

```
Qy        1060 ATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAAC 1119
                |    |      || || |  | |   | ||  |  | || || |||   ||| |  ||
Db         466 TTCTCAGACAGTTTGCAAGAAATAATGAAATCTATTCTTACTTTACAATCCCAATTAGAC 525

Qy        1120 TCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGA 1179
               ||  | ||||||| |    || ||||| ||  |||    | || | || ||  ||  | | |
Db         526 TCTTTGGCAGCAATGACTCTCCAAAACCGCCGAGGCCCACACCTCCTCACTGCTGAGAAA 585

Qy        1180 GGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATT 1239
               || |||  |||    || |||| ||||| | |||| |||   ||| || ||  | ||
Db         586 GGAGGACTCTGCACCTTCTTAGGGGAAGAGTGTTGTTTTTACACTAACCAGTCAGGGATA 645

Qy        1240 GTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAA 1299
               ||    ||           |  | |  ||   |          ||    | |    |||  |
Db         646 GTACGAGATGCCACCTGGCATTTACAGGAAAGGGCTTCTGATATCAGACAATGCCTTTCA 705

Qy        1300 AACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGA 1359
               ||| |  |   |        |||| |    |||    | ||| |||| ||  ||  ||||
Db         706 AACTCTTATACCAA------CCTCTGGAGTTGGGCAACATGGCTTCTTCCATTTCTAGGT 759

Qy        1360 CCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTT 1419
               ||  | |||||  |  |  ||||||||  |||||| |||||||| ||||| || |||||
Db         760 CCCATGGCAGCCATCTTGCTGTTACTCACCTTTGGGCCCTGTATTTTTAAGCTTCTTGTC 819

Qy        1420 AAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCC 1479
               || ||||| || || || || |||||  | ||||||||||||||||||||||||||||||
Db         820 AAATTTGTTTCCTCTAGGATCGAAGCCATCAAGCTACAGATGGTCTTACAAATGGAACCC 879

Qy        1480 CA 1481
               ||
Db         880 CA 881


RESULT 7
US-09-374-766-58
; Sequence 58, Application US/09374766
; Patent No. 6579526
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/374,766
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;     TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 base pairs
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-374-766-58

  Query Match             9.8%;  Score 145.6;  DB 4;  Length 1722;
  Best Local Similarity   55.7%;  Pred. No. 1.1e-36;
  Matches  302;  Conservative    0;  Mismatches  234;  Indels    6;  Gaps    1;

Qy        940 CACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTA 999
              ||   || |    |||   |  |  |  |  ||||  |   |||| ||           |
Db        346 CAGTGCAGAGCCATACAACTAATATCCCTATTTATAGGGTTAGGAATGGCTACTGCTACA 405

Qy       1000 GGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAA 1059
              || ||||| || | | || |  | || ||       | |||| | | ||| ||  |  |
Db        406 GGAACTGGAATAGCCGGTTTATCTACTTCATTATCCTACTACCATACACTCTCAAAGAAT 465

Qy       1060 ATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAAC 1119
               |    |      || || |  | |   | ||  |  | || || |||   ||| |  ||
Db        466 TTCTCAGACAGTTTGCAAGAAATAATGAAATCTATTCTTACTTTACAATCCCAATTAGAC 525

Qy       1120 TCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGA 1179
              ||  | |||||| |    || ||||| ||  |||    | || | || || ||  | | |
Db        526 TCTTTGGCAGCAATGACTCTCCAAAACCGCCGAGGCCCACACCTCCTCACTGCTGAGAAA 585

Qy       1180 GGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATT 1239
              || |||  |||    || ||||| |||||  | |||| |||   ||| || ||  | ||
Db        586 GGAGGACTCTGCACCTTCTTAGGGGAAGAGTGTTGTTTTTACACTAACCAGTCAGGGATA 645

Qy       1240 GTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAA 1299
              ||    ||           |  | |  ||  |         ||    | |    |||  |
Db        646 GTACGAGATGCCACCTGGCATTTACAGGAAAGGGCTTCTGATATCAGACAATGCCTTTCA 705

Qy       1300 AACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGA 1359
              ||| |  |   |        |||| |    |||    | ||| |||| || ||  ||||
Db        706 AACTCTTATACCAA------CCTCTGGAGTTGGGCAACATGGCTTCTTCCATTTCTAGGT 759

Qy       1360 CCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTT 1419
              ||  | |||||  |  |  ||||||||  |||||| |||||||| ||||| || |||||
Db        760 CCCATGGCAGCCATCTTGCTGTTACTCACCTTTGGGCCCTGTATTTTTAAGCTTCTTGTC 819

Qy       1420 AAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCC 1479
              || ||||| || || || || |||||  | ||||||||||||||||||||||||||||||
Db        820 AAATTTGTTTCCTCTAGGATCGAAGCCATCAAGCTACAGATGGTCTTACAAATGGAACCC 879

Qy       1480 CA 1481
              ||
Db        880 CA 881


RESULT 8
US-08-979-847B-54
; Sequence 54, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,847B
;            FILING DATE: 26-No. 6582703-1997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;    INFORMATION FOR SEQ ID NO: 54:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1722 base pairs
;            TYPE: nucleotide
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-08-979-847B-54

  Query Match             9.8%;  Score 145.6;  DB 4;  Length 1722;
  Best Local Similarity   55.7%;  Pred. No. 1.1e-36;
  Matches  302;  Conservative    0;  Mismatches  234;  Indels    6;  Gaps    1;

Qy        940 CACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTA 999
              ||   || |    |||   |  |  |  |  ||||  |   |||| ||           |
Db        346 CAGTGCAGAGCCATACAACTAATATCCCTATTTATAGGGTTAGGAATGGCTACTGCTACA 405

Qy       1000 GGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAA 1059
              || ||||| || | | || |  | || ||       | |||| | | ||| ||  |  |
Db        406 GGAACTGGAATAGCCGGTTTATCTACTTCATTATCCTACTACCATACACTCTCAAAGAAT 465

Qy       1060 ATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAAC 1119
               |    |      || || |  | |   | ||  |  | || || |||   ||| |  ||
Db        466 TTCTCAGACAGTTTGCAAGAAATAATGAAATCTATTCTTACTTTACAATCCCAATTAGAC 525

Qy       1120 TCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGA 1179
              ||  | |||||| |    || ||||| ||  |||    | || | || || ||  | | |
Db        526 TCTTTGGCAGCAATGACTCTCCAAAACCGCCGAGGCCCACACCTCCTCACTGCTGAGAAA 585

Qy       1180 GGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATT 1239
              || |||  |||    || ||||| |||||  | |||| |||   ||| || ||  | ||
Db        586 GGAGGACTCTGCACCTTCTTAGGGGAAGAGTGTTGTTTTTACACTAACCAGTCAGGGATA 645

Qy       1240 GTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAA 1299
              ||    ||           |  | |  ||  |         ||    | |    |||  |
Db        646 GTACGAGATGCCACCTGGCATTTACAGGAAAGGGCTTCTGATATCAGACAATGCCTTTCA 705

Qy       1300 AACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGA 1359
              ||| |  |   |        |||| |    |||    | ||| |||| || ||  ||||
Db        706 AACTCTTATACCAA------CCTCTGGAGTTGGGCAACATGGCTTCTTCCATTTCTAGGT 759

Qy       1360 CCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTT 1419
              ||  | |||||  |  |  ||||||||  |||||| |||||||| ||||| || |||||
Db        760 CCCATGGCAGCCATCTTGCTGTTACTCACCTTTGGGCCCTGTATTTTTAAGCTTCTTGTC 819

Qy       1420 AAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCC 1479
              || ||||| || || || || |||||  | ||||||||||||||||||||||||||||||
Db        820 AAATTTGTTTCCTCTAGGATCGAAGCCATCAAGCTACAGATGGTCTTACAAATGGAACCC 879
```

```
Qy       1480 CA 1481
              ||
Db        880 CA 881


RESULT 9
US-09-799-451-249
; Sequence 249, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Goodrich, Ryle
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Yamazaki, Victoria
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Ghosh, Reena
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6783969el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 803
;  CURRENT APPLICATION NUMBER: US/09/799,451
;  CURRENT FILING DATE:  2001-03-05
;  NUMBER OF SEQ ID NOS: 948
;  SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 249
;   LENGTH: 2908
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (382)..(1995)
US-09-799-451-249

  Query Match             9.8%;  Score 144.6;  DB 4;  Length 2908;
  Best Local Similarity   56.9%;  Pred. No. 3.2e-36;
  Matches  286;  Conservative    0;  Mismatches  214;  Indels    3;  Gaps    1;

Qy        962 TTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCA 1021
              ||||  || |     ||   ||  | ||||   |    || || || ||||   | ||||
Db       1445 TTCCCCTTCTCGCGGGACTCGGCATTCTAGCTGGTACGGGAACCGGAATTGCTGGAATCA 1504

Qy       1022 CAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGG 1081
              |||   || | |    |||   | | || ||  | ||||||       |||| ||
Db       1505 CAAAAGCTTCCCTCACCTATAGCCAGCTCTCAAAGGAAATAGCCAACAACATTGACACCA 1564

Qy       1082 TCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTC 1141
              |  || |  || |    ||| ||||||| ||| |  ||||  ||||||| ||||||||||
Db       1565 TGGCTAAAGCCTTAACGACCATGCAAGAACAAATCGACTCTTTAGCAGCCGTAGTCCTTC 1624

Qy       1142 AAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAG 1201
              |||||||  |||   ||||| || |||| ||   |   || ||||  |||||    ||||
Db       1625 AAAATCGTCGAGGACTAGACATGTTAACGGCAGCACAGGGAGGAATTTGTTTGGCCTTAG 1684

Qy       1202 GAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAA 1261
                ||| || | || | ||  || ||||||||  |||  ||    || ||  | | | ||
Db       1685 ATGAAAAATGTTGCTTTTGGGTAAATCAATCAGGAAAAGTACAAGACAACATCAGACAAC 1744

Qy       1262 TTCGAGATCGAATACAATGT---AGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCC 1318
              | | | ||| |       ||    | | |   |    |||         |   ||||
Db       1745 TCCTAAATCAAGCCTCCAGTTTACGGGAACGAGCCACTCAGGGTTGGTTAAATTGGGAAG 1804

Qy       1319 TCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATAT 1378
```

```
                    |  ||||| |  | |||||||| ||| |   ||| || || |    |||  ||
Db       1805 GAACTTGGAAATGGTTCTCTTGGGTTCTTCCCCTTACAGGCCCACTTGTTAGTCTCCTAC 1864

Qy       1379 TGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAA 1438
              | || ||||| ||||| || ||| || | || ||  |     | |||||||| ||  |
Db       1865 TTTTGCTCCTTTTTGGTCCATGTCTCCTAAATCTAATAACCCAATTTGTCTCCTCTCGCC 1924

Qy       1439 TTGAAGCTGTAAAGCTACAGATG 1461
              || | ||  ||||||| |||| |
Db       1925 TTCAGGCCATAAAGCTCCAGACG 1947


RESULT 10
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTI
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563C
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-691-563C-46

  Query Match             8.9%;  Score 131.2;  DB 3;  Length 1859;
  Best Local Similarity   57.6%;  Pred. No. 5.6e-32;
  Matches  257;  Conservative    0;  Mismatches  183;  Indels    6;  Gaps    1;

Qy       1036 TTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTG 1095
              | |||| ||| ||| ||  | ||  |    |      || || |  | |     ||  |
Db       1092 TACTACCACACACTCTCAAAGGATTTCTCAGACAGTTTGCAAGAAATAATGATATCTATC 1151

Qy       1096 GTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCT 1155
               | ||  | |||   ||| |  ||||  | ||||||||    || ||||| ||   ||
Db       1152 CTTACTCTACAATCCCAAATAGACTCTTTGGCAGCAGTGACTCTCCAAAACCGTCAAGGC 1211

Qy       1156 TTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGT 1215
               ||||| | || || ||  | | ||| |||  |||    || ||| | |||||  | |||
Db        1212 CTAGACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTAAGGGAAGAGTGTTGT 1271

Qy        1216 TATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATA 1275
                 |||   ||| || ||  | || ||    ||    |      |  | |   |
Db        1272 CTTTACACTAACCAGTCAGGGATAGTATGAGATGCTGCCCGGCATTTACAGAAAAAGGCT 1331

Qy        1276 CAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATG 1335
                     |    | |    |||  |||  |      ||    |  |||| |    |||
Db        1332 TCTGAAATCAGACAACGCCTTTCAAATTC------CTATACCAACCTCTGGAGTTGGGCA 1385

Qy        1336 CCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGA 1395
                | ||| |||| |||||  || | ||  | || ||  |  |  | ||||||  ||||||
Db        1386 ACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGGG 1445

Qy        1396 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1455
               |||||||| |||||||||||||| || ||||| ||||| || || || ||  | ||||||
Db        1446 CCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCTA 1505

Qy        1456 CAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||||||||||||||||||
Db        1506 CAGATGGTCTTACAAATGGAACCCCA 1531


RESULT 11
US-09-374-766-46
; Sequence 46, Application US/09374766
; Patent No. 6579526
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/374,766
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1859 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-374-766-46
```

```
  Query Match             8.9%;  Score 131.2;  DB 4;  Length 1859;
  Best Local Similarity   57.6%;  Pred. No. 5.6e-32;
  Matches  257;  Conservative    0;  Mismatches  183;  Indels    6;  Gaps    1;

Qy        1036 TTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTG 1095
               | |||| ||| ||| ||  | ||  |    |      || || |  | |     ||  |
Db        1092 TACTACCACACACTCTCAAAGGATTTCTCAGACAGTTTGCAAGAAATAATGATATCTATC 1151

Qy        1096 GTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCT 1155
                | ||  | |||   ||| |  ||||  | ||||||||    || ||||| ||   ||
Db        1152 CTTACTCTACAATCCCAAATAGACTCTTTGGCAGCAGTGACTCTCCAAAACCGTCAAGGC 1211

Qy        1156 TTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGT 1215
                ||||| | || || ||  | | ||| |||  |||    || ||| | |||||  | |||
Db        1212 CTAGACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTAAGGGAAGAGTGTTGT 1271

Qy        1216 TATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATA 1275
                 |||   ||| || ||  | || ||    ||    |      |  | |   |
Db        1272 CTTTACACTAACCAGTCAGGGATAGTATGAGATGCTGCCCGGCATTTACAGAAAAAGGCT 1331

Qy        1276 CAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATG 1335
                     |    | |    |||  |||  |      ||    |  |||| |    |||
Db        1332 TCTGAAATCAGACAACGCCTTTCAAATTC------CTATACCAACCTCTGGAGTTGGGCA 1385

Qy        1336 CCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGA 1395
                | ||| |||| |||||  || | ||  | || ||  |  |  | ||||||  ||||||
Db        1386 ACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGGG 1445

Qy        1396 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1455
               |||||||| |||||||||||||| || ||||| ||||| || || || ||  | ||||||
Db        1446 CCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCTA 1505

Qy        1456 CAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||||||||||||||||||
Db        1506 CAGATGGTCTTACAAATGGAACCCCA 1531


RESULT 12
US-08-979-847B-42
; Sequence 42, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
;         APPLICANT: PERRON, HERVE
;                    BESEME, FREDERIC
;                    BEDIN, FREDERIC
;                    PARANHOS-BACCALA, GLAUCIA
;                    KOMURIAN-PRADEL, FLORENCE
;                    JOLIVET-REYNAUD, COLETTE
;                    MANDRAND, BERNARD
;                    GARSON, JEREMY
;                    TUKE, PHILIP
;         TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                             ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                             THERAPEUTIC PURPOSES
;         NUMBER OF SEQUENCES: 210
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: OLIFF & BERRIDGE, PLC
;              STREET: P.O. BOX 19928
;              CITY: ALEXANDRIA
;              STATE: VA
;              COUNTRY: USA
;              ZIP: 22320
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/08/979,847B
;              FILING DATE: 26-No. 6582703-1997
;              CLASSIFICATION: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: BERRIDGE, WILLIAM P.
;              REGISTRATION NUMBER: 30,024
;              REFERENCE/DOCKET NUMBER: WPB 39046A
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 703-836-6400
;              TELEFAX: 703-836-2787
;    INFORMATION FOR SEQ ID NO: 42:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1859 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-08-979-847B-42

  Query Match             8.9%;  Score 131.2;  DB 4;  Length 1859;
  Best Local Similarity   57.6%;  Pred. No. 5.6e-32;
  Matches  257;  Conservative    0;  Mismatches  183;  Indels    6;  Gaps    1;

Qy       1036 TTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTG 1095
              | |||| ||| ||| ||  | ||  |    |      || || |  | |     ||  |
Db       1092 TACTACCACACACTCTCAAAGGATTTCTCAGACAGTTTGCAAGAAATAATGATATCTATC 1151

Qy       1096 GTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCT 1155
               | ||  | |||   ||| |  ||||  | ||||||||    || ||||| ||   ||
Db       1152 CTTACTCTACAATCCCAAATAGACTCTTTGGCAGCAGTGACTCTCCAAAACCGTCAAGGC 1211

Qy       1156 TTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGT 1215
               ||||| | || || ||  | | ||| |||  |||    || ||| | |||||  | |||
Db       1212 CTAGACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTAAGGGAAGAGTGTTGT 1271

Qy       1216 TATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATA 1275
                |||   ||| || ||  | || ||    ||    |      |  | |   |
Db       1272 CTTTACACTAACCAGTCAGGGATAGTATGAGATGCTGCCCGGCATTTACAGAAAAAGGCT 1331

Qy       1276 CAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATG 1335
                    |    | |    |||  |||  |      ||    |  |||| |    |||
Db       1332 TCTGAAATCAGACAACGCCTTTCAAATTC------CTATACCAACCTCTGGAGTTGGGCA 1385

Qy       1336 CCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGA 1395
               | ||| |||| |||||  || | ||  | || ||  |  |  | ||||||  ||||||
Db       1386 ACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTATTACTCGCCTTTGGG 1445

Qy       1396 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1455
              |||||||| |||||||||||||| || ||||| ||||| || || || ||  | ||||||
Db       1446 CCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCTA 1505

Qy       1456 CAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||||||||||||||||||
Db       1506 CAGATGGTCTTACAAATGGAACCCCA 1531

RESULT 13
US-09-671-317-406
; Sequence 406, Application US/09671317
; Patent No. 6528260
; GENERAL INFORMATION:
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Cohen, Annick
;  TITLE OF INVENTION:  BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
;  FILE REFERENCE:  62.US3.CIP
;  CURRENT APPLICATION NUMBER: US/09/671,317
;  CURRENT FILING DATE:  2000-09-27
;  PRIOR APPLICATION NUMBER:  US 09/536,178
;  PRIOR FILING DATE:  2000-03-23
;  PRIOR APPLICATION NUMBER:  PCT/IB00/00403
;  PRIOR FILING DATE:  2000-03-24
;  PRIOR APPLICATION NUMBER:  US 60/126,269
;  PRIOR FILING DATE:  1999-03-25
;  PRIOR APPLICATION NUMBER:  US 60/131,961
;  PRIOR FILING DATE:  1999-04-30
```

```
;  NUMBER OF SEQ ID NOS:  977
;  SOFTWARE:  Patent.pm
; SEQ ID NO 406
;   LENGTH: 1001
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 503
;   OTHER INFORMATION: 12-5-158 : polymorphic base C or T
;   NAME/KEY: misc_binding
;   LOCATION: 484..502
;   OTHER INFORMATION: 12-5-158.mis1
;   NAME/KEY: misc_binding
;   LOCATION: 504..523
;   OTHER INFORMATION: 12-5-158.mis2, potential complement
;   NAME/KEY: primer_bind
;   LOCATION: 346..366
;   OTHER INFORMATION: upstream amplification primer
;   NAME/KEY: primer_bind
;   LOCATION: 801..821
;   OTHER INFORMATION: downstream amplification primer, complement
;   NAME/KEY: misc_binding
;   LOCATION: 491..515
;   OTHER INFORMATION: 12-5-158 potential probe
;   NAME/KEY: misc_feature
;   LOCATION: 314,336,793
;   OTHER INFORMATION: n=a, g, c or t
US-09-671-317-406

  Query Match             8.8%;  Score 130.2;  DB 4;  Length 1001;
  Best Local Similarity   67.6%;  Pred. No. 8e-32;
  Matches  211;  Conservative    0;  Mismatches   99;  Indels    2;  Gaps    2;

Qy        773 CCACACGAATAGTCTGCCTACCCTCAGG-AATATTTTTTGTCTGTGGTACCTCAGCCTAT 831
              || || |  |   ||  ||||| ||||| ||| |||||  |||||| || | ||||||||
Db        688 CCTCAGGTTTCACCTATCTACCTTCAGGAAATTTTTTTCATCTGTGATAACACAGCCTAT 747

Qy        832 CATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCT-TAGTGCCCCCTAT 890
              |  ||    |||||| || || ||    | ||||   |||||||   |||  || || ||
Db        748 CGATGCCAAAATGGCACTCCAAAAGAACTATGCTATGTCTCATTTNCTAGCACCTCCCAT 807

Qy        891 GACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAG 950
              | |||| || ||||||||||| ||| | | ||   |  ||||  |  | | | |||   |
Db        808 GTCCATATATACTGAACAAGAGTTACAAAGTCTCCTTATACCCCAATCTCGCCACACCCG 867

Qy        951 AGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCAT 1010
              ||  |  ||| | ||||||    |  |||||||| | |||||||| || || ||||| ||
Db        868 AGCCCTTATTGTCCCTTTTACAGTAGGAGCAGGAATACTAGGCAGGCTTGGGACTGGAAT 927

Qy       1011 TGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGA 1070
              |||  | || ||  |||| || || ||||| || ||| ||||   |||| |||||| |||
Db        928 TGGAGGCATAACCTCCTCCACCCAATTCTATTATAAATTATCATGAGAATTAAATGATGA 987

Qy       1071 CATGGAACAGGT 1082
              |||||||   ||
Db        988 CATGGAATGAGT 999

RESULT 14
US-08-007-282B-1
; Sequence 1, Application US/08007282B
; Patent No. 5403582
;  GENERAL INFORMATION:
;     APPLICANT:  NAZERIAN, KEYVAN
;     APPLICANT:  CALVERT, JAY G.
;     APPLICANT:  WITTER, RICHARD L.
;     APPLICANT:  YANAGIDA, NOBORU
;     TITLE OF INVENTION:  VACCINE COMPRISING FOWLPOX VIRUS
;     TITLE OF INVENTION:  RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF AN
;     TITLE OF INVENTION:  AVIAN RETICULOENDOTHELIOSIS RETROVIRUS
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  8110 Gatehouse Road  Suite 500 East
;      CITY:  Falls Church
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/007,282B
;      FILING DATE:  19930121
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY JR., GERALD M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1644-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1704 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-007-282B-1

  Query Match              7.9%;  Score 116.8;  DB 1;  Length 1704;
  Best Local Similarity   51.5%;  Pred. No. 2.6e-27;
  Matches  268;  Conservative    0;  Mismatches  252;  Indels    0;  Gaps    0;

Qy        932 CTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAG 991
              ||| ||  || || |    ||| |   || | ||  |  || |  |   |||  |   ||
Db       1178 CTAGGCGTCATAAGAGGGCAGTCCAGTTTATCCCCCTGCTTGTGGGTCTAGGGATTTCAG 1237

Qy        992 GCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTAT 1051
              |      |  | ||||        |  |      |||   |||    ||  |||| || |
Db       1238 GGGCTACACTTGCTGGTGGAACGGGGCTTGGGGTCTCCGTTCACACTTATCACAAGCTCT 1297

Qy       1052 CTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATC 1111
              || |  || | | ||  ||  |  |     |  | |   || |      | | || || |
Db       1298 CTAATCAATTGATTGAAGATGTCCAGGCTCTTTCAGGGACCATCAATGACCTACAGGACC 1357

Qy       1112 AACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCG 1171
              |  || ||||||| || |  || ||| | ||||||| ||||||  |||||| |  | || |
Db       1358 AGATTGACTCCCTGGCTGAGGTTGTCTTACAAAATAGAAGAGGGTTAGACCTATTGACTG 1417

Qy       1172 CCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAAT 1231
              || ||  ||| ||||  ||| |     |    ||  |  | |||| ||| | |||  | |
Db       1418 CCGAACAAGGAGGAATATGTCTCGCACTCCAGGAGAAGTGTTGTTTTTACGCTAACAAGT 1477

Qy       1232 CCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGG 1291
              |  | || ||   ||| ||  |   | || | | |||     |       |||  |   |
Db       1478 CGGGTATCGTACGTGACAAGATCCGAAAACTCCAAGAGGACCTTATCGAGAGAAAACGTG 1537

Qy       1292 AGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCT 1351
                ||  |  ||| |   |  ||| ||  | | | |   |   | ||  |  | || ||||
Db       1538 CACTGTACGACAACCCCCTGTGGAGCGGCTTGAACGGCTTCCTTCCATATTTGCTACCCT 1597

Qy       1352 TCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACC 1411
              | ||||| || ||    |  || ||||||||| ||   | | || || || ||  | ||
Db       1598 TGTTAGGCCCCCTGTTTGGGCTCATATTGTTCCTGACCCTCGGCCCGTGCATTATGAAGA 1657

Qy       1412 TCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAA 1451
               |||    |   || |   |  || |||| | || |||||
Db       1658 CCCTGACTCGCATTATACATGACAAAATTCAGGCAGTAAA 1697
```

```
RESULT 15
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

  Query Match              7.8%;  Score 114.8;  DB 3;  Length 80246;
  Best Local Similarity   55.7%;  Pred. No. 1.3e-25;
  Matches  263;  Conservative    0;  Mismatches  202;  Indels    7;  Gaps    2;

Qy       1010 TTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTG 1069
              ||  |||  | |||     ||  |   |     |||  || ||  | ||  |    |
Db      58509 TTAACAGCGTAACATGTATTATCCTACTACCTCACACTCTCTCAAAGGATTTCTCAGACA 58568

Qy       1070 ACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAG 1129
                 || || |    ||  | ||  |  | ||  | |||   ||| |  ||||  | ||||
Db      58569 GTTTGCAAAAAAGAACGAAATCTGTCCTTACTCTACAATCCCAAATAGACTCTTTGGCAG 58628

Qy       1130 CAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCT 1189
              ||||    || ||||| ||  |||   ||||||  || ||  ||  | | |||  ||  ||
Db      58629 CAGTGACTCTCCAAAACCGCTGAGGCCTAGACTCTCTTACTGCTGAGAAAGGAAGATTCT 58688

Qy       1190 GTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGA 1249
              |    || ||||| | |||  | |||| ||||  ||| || ||  | ||  |    || |
Db      58689 GCAC-TTCTTAGGGGTAGAGTGTTGTTTTTATACTAACCAGTCAGGGATAATATGAGATA 58747

Qy       1250 AAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAAC 1309
                    |     | |  ||               |    | |    |||  |||| |  |
Db      58748 CCACCCAGTGTTTACAGGAAAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCTTATA 58807

Qy       1310 GCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAG 1369
               |        |||| |    |||  | | ||| ||||||||||  |||| ||| |  |||
Db      58808 CCAA------CCTCTGGAGTTGGGCGACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAG 58861

Qy       1370 CTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCT 1429
              |  |  |  |  || ||   ||||| |||||||| ||||||||| | || || ||||| |
Db      58862 CCATCTTGCTAATAGTCGCATTTGGGCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTT 58921

Qy       1430 CTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              | || || || || ||  | |||||||||||| ||||||||||| |||||||
Db      58922 CCTCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCA 58973


RESULT 16
US-09-011-745-1
; Sequence 1, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2518
;   TYPE: DNA
;   ORGANISM: RD114
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-1

  Query Match             7.7%;  Score 114.2;  DB 3;  Length 2518;
  Best Local Similarity   55.0%;  Pred. No. 2.3e-26;
  Matches  224;  Conservative    0;  Mismatches  183;  Indels    0;  Gaps    0;

Qy       1045 AAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTG 1104
              ||| |||| ||  |  |||    |||  |  ||    |  | |   || |       ||
Db       1834 AAATTATCCCATCAGTTAATATCTGATGTCCAAGTCTTATCCGGTACCATACAAGATTTA 1893

Qy       1105 CAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTG 1164
              ||||| ||  |  ||||  |||| | |||||| || |||||| | || |   | ||| |
Db       1894 CAAGACCAGGTAGACTCGTTAGCTGAAGTAGTTCTCCAAAATAGGAGGGGACTGGACCTA 1953

Qy       1165 CTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTT 1224
              ||||| ||  ||  ||| ||||  ||||||   |||  |||| || |||||| ||||| |
Db       1954 CTAACGGCAGAACAAGGAGGAATTTGTTTAGCCTTACAAGAAAAATGCTGTTTTTATGCT 2013

Qy       1225 AATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGA 1284
              ||  | ||  ||||||| |   | ||| | | |    | | |||   | |||||    |
Db       2014 AACAAGTCAGGAATTGTGAGAAACAAAATAAGAACCCTACAAGAAGAATTACAAAAACGC 2073

Qy       1285 GCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTT 1344
                 ||    ||   || || |   | ||||  |   ||       |   | || |   |
Db       2074 AGGGAAAGCCTGGCAACCAACCCTCTCTGGACCGGGCTGCAGGGCTTTCTTCCGTACCTC 2133

Qy       1345 CTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATC 1404
              || ||  || | ||||| |||    | ||  || |  ||||   | |||| || ||  |
Db       2134 CTACCTCTCCTGGGACCCCTACTCACCCTCCTACTCATACTAACCATTGGGCCATGCGTT 2193

Qy       1405 TTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAA 1451
              || |  | |||  |    ||  |   |   ||| ||  | | |||| |
Db       2194 TTCAGTCGCCTCATGGCCTTCATTAATGATAGACTTAATGTTGTACA 2240

RESULT 17
US-09-011-745-8
; Sequence 8, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 5865
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
```

```
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3611)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3612)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3613)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3614)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3799)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3800)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3801)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3802)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-8

  Query Match             7.7%;  Score 114.2;  DB 3;  Length 5865;
  Best Local Similarity   55.0%;  Pred. No. 3.9e-26;
  Matches  224;  Conservative    0;  Mismatches  183;  Indels    0;  Gaps    0;

Qy       1045 AAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTG 1104
              ||| |||| ||  |  |||    |||  |  ||    |  | |   || |       ||
Db       2692 AAATTATCCCATCAGTTAATATCTGATGTCCAAGTCTTATCCGGTACCATACAAGATTTA 2751

Qy       1105 CAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTG 1164
              ||||| ||  |  ||||  |||| | |||||| || |||||| | || |   | ||| |
Db       2752 CAAGACCAGGTAGACTCGTTAGCTGAAGTAGTTCTCCAAAATAGGAGGGGACTGGACCTA 2811

Qy       1165 CTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTT 1224
              ||||| ||  ||  ||| ||||  ||||||   |||  |||| || |||||| ||||| |
Db       2812 CTAACGGCAGAACAAGGAGGAATTTGTTTAGCCTTACAAGAAAAATGCTGTTTTTATGCT 2871

Qy       1225 AATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGA 1284
              ||  | ||  ||||||| |   | ||| | | |    | | |||   | |||||    |
Db       2872 AACAAGTCAGGAATTGTGAGAAACAAAATAAGAACCCTACAAGAAGAATTACAAAAACGC 2931

Qy       1285 GCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTT 1344
                 ||    ||   || || |   | ||||  |   ||       |   | || |   |
Db       2932 AGGGAAAGCCTGGCAACCAACCCTCTCTGGACCGGGCTGCAGGGCTTTCTTCCGTACCTC 2991

Qy       1345 CTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATC 1404
              || ||  || | ||||| |||    | ||  || |  ||||   | |||| || ||  |
Db       2992 CTACCTCTCCTGGGACCCCTACTCACCCTCCTACTCATACTAACCATTGGGCCATGCGTT 3051

Qy       1405 TTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAA 1451
              || |  | |||  |    ||  |   |   ||| ||  | | |||| |
Db       3052 TTCAGTCGCCTCATGGCCTTCATTAATGATAGACTTAATGTTGTACA 3098

RESULT 18
US-09-555-352-1
; Sequence 1, Application US/09555352
; Patent No. 6544779
; GENERAL INFORMATION:
```

```
; APPLICANT: Cichutek, Klaus
; APPLICANT:  Merget-Millitzer, Heike
; TITLE OF INVENTION: PSEUDO-TYPE RETROVIRAL VECTORS WITH
; TITLE OF INVENTION:  MODIFIABLE SURFACE CAPSID PROTEINS
; FILE REFERENCE: 11692-005001
; CURRENT APPLICATION NUMBER: US/09/555,352
; CURRENT FILING DATE:  2000-08-24
; PRIOR APPLICATION NUMBER: PCT/DE98/03542
; PRIOR FILING DATE: 1998-11-27
; PRIOR APPLICATION NUMBER: DE 197 52 855.4
; PRIOR FILING DATE: 1997-11-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 4776
;   TYPE: DNA
;   ORGANISM: Murine leukemia virus
US-09-555-352-1

  Query Match              7.6%;  Score 112.2;  DB 4;  Length 4776;
  Best Local Similarity   51.5%;  Pred. No. 1.5e-25;
  Matches  258;  Conservative    0;  Mismatches  243;  Indels    0;  Gaps    0;

Qy        951 AGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCAT 1010
              ||| |   || | ||  |  || |  |   |||  |   |||      |  | ||||
Db       1216 AGTCCAGTTTATCCCCCTGCTTGTGGGTCTAGGGATTTCAGGGGCTACACTTGCTGGTGG 1275

Qy       1011 TGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGA 1070
                   |  |      |||   |||    ||  |||| || ||| |  || | | ||  ||
Db       1276 AACGGGGCTTGGGGTCTCCGTTCACACTTATCACAAGCTCTCTAATCAATTGATTGAAGA 1335

Qy       1071 CATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGC 1130
                |  |     |  | |   || |      | | || || ||  || ||||||| || |
Db       1336 TGTCCAGGCTCTTTCAGGGACCATCAATGACCTACAGGACCAGATTGACTCCCTGGCTGA 1395

Qy       1131 AGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTG 1190
               || ||| | |||||| ||||||  |||||| |  | || ||| ||  ||| ||||  ||
Db       1396 GGTTGTCTTACAAAATAGAAGAGGGTTAGACCTATTGACTGCCGAACAAGGAGGAATATG 1455

Qy       1191 TTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAA 1250
              | |     |    | | |  | |||| ||| | |||  |  ||  | || ||   ||| ||
Db       1456 TCTCGCACTCCAGGAGAAGTGTTGTTTTTACGCTAACAAGTCGGGTATCGTACGTGACAA 1515

Qy       1251 AGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACG 1310
                |   | || | | |||     |       |||  |   |  ||  |  ||| |   |
Db       1516 GATCCGAAAACTCCAAGAGGACCTTATCGAGAGAAAACGTGCACTGTACGACAACCCCCT 1575

Qy       1311 CTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGC 1370
               ||| ||  | | | |   |   | || |   | || ||||| ||||| || ||    |
Db       1576 GTGGAGCGGCTTGAACGGCTTCCTTCCATATTTGCTACCCTTGTTAGGCCCCCTGTTTGG 1635

Qy       1371 TCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTC 1430
               || |||||||| ||   | | || || || ||  | ||   |||   |    || |
Db       1636 GCTCATATTGTTCCTGACCCTCGGCCCGTGCATTATGAAGACCCTGACTCGCATTATACA 1695

Qy       1431 TTCCAGAATTGAAGCTGTAAA 1451
              |  || |||| | || |||||
Db       1696 TGACAAAATTCAGGCAGTAAA 1716

RESULT 19
US-09-620-312D-171
; Sequence 171, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
```

```
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  John Tillinghast
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6569662el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/09/620,312D
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1105
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 171
;   LENGTH: 2809
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1936)..(2364)
US-09-620-312D-171

  Query Match              7.2%;  Score 106.4;  DB 4;  Length 2809;
  Best Local Similarity   55.0%;  Pred. No. 8.6e-24;
  Matches  231;  Conservative    0;  Mismatches  186;  Indels    3;  Gaps    1;

Qy        996 ACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCA 1055
              ||| || ||||| || |   | || || ||||| |     |  ||| | | |||||
Db       1851 ACTTGGCACTGGAATAGCAGGCATAACCACCTCAATCACCTCATACCAAACACTATTCAC 1910

Qy       1056 AGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACT 1115
              |    |   |     | | ||| |  |       ||| |   ||   | |||   ||| |
Db       1911 AACCCTTTCTAACACCGTAGAAGATATGCACACTTCCATTACCAGTCTCCAACGACAATT 1970

Qy       1116 TAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAA 1175
                ||| ||| |  | |||  ||||||||||  | ||||   | ||| | |||||| |  |
Db       1971 AGACTTCCTCGTGGGAGTCATCCTTCAAAACTGGAGAGTCCTGGACCTCCTAACCACTGA 2030

Qy       1176 AAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAG 1235
               | |||||| |||||  |||   |    |||||| ||||||  | ||||||| ||||  |
Db       2031 GAAAGGGGGTACCTGCATATACCTCCAGGAAGAATGCTGTTTCTGTGTTAATGAATCTGG 2090

Qy       1236 AATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCG---AATACAATGTAGAGCAGAGGA 1292
               |||||   |     ||||       |||  ||  |      | |   ||||    | |
Db       2091 CATTGTTCATATCGCAGTTCGTAGGCTTCATGACAGGGCTGCAGAGCTTAGACATCAAGT 2150

Qy       1293 GCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTT 1352
                 | |   |      |   |      |||      |||||| |||||||||  ||||||
Db       2151 CGCTGACTCCTGGTGGCAAGGATCATCCCTTCTAAGATGGATACCCTGGGTTGCCCCCTT 2210

Qy       1353 CTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCT 1412
              | ||||||| ||       ||  |  |||||||  |  |||| || || || ||||||||
Db       2211 CCTAGGACCCCTGATCTTCCTCTTCCTGTTACTAATGATTGGGCCATGCATATTTAACCT 2270

RESULT 20
US-09-904-615-12
; Sequence 12, Application US/09904615
; Patent No. 6566325
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 49 Human Secreted Proteins
;  FILE REFERENCE: PZ032P1
;  CURRENT APPLICATION NUMBER: US/09/904,615
;  CURRENT FILING DATE:  2001-07-16
;  PRIOR APPLICATION NUMBER: 09/511,554
;  PRIOR FILING DATE: 2000-02-23
;  PRIOR APPLICATION NUMBER: 60/097,917
;  PRIOR FILING DATE: 1998-08-25
;  PRIOR APPLICATION NUMBER: 60/098,634
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 2342
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-904-615-12

  Query Match              7.1%;  Score 104.8;  DB 4;  Length 2342;
  Best Local Similarity   54.8%;  Pred. No. 2.5e-23;
  Matches  230;  Conservative    0;  Mismatches  187;  Indels    3;  Gaps    1;

Qy         996 ACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCA 1055
               ||| || ||||| || |   | || || ||||| |     |  ||| | | |||||
Db        1357 ACTTGGCACTGGAATAGCAGGCATAACCACCTCAATCACCTCATACCAAACACTATTCAC 1416

Qy        1056 AGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACT 1115
               |    |   |     | | ||| |  |       ||| |   ||   | |||   ||| |
Db        1417 AACCCTTTCTAACACCGTAGAAGATATGCACACTTCCATTACCAGTCTCCAACGACAATT 1476

Qy        1116 TAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAA 1175
                 ||| ||| |  | |||  ||||||||||  | ||||   | ||| | |||||| |  |
Db        1477 AGACTTCCTCGTGGGAGTCATCCTTCAAAACTGGAGAGTCCTGGACCTCCTAACCACTGA 1536

Qy        1176 AAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAG 1235
                | |||||| |||||  |||   |    |||||| ||||||  | ||||||| ||||  |
Db        1537 GAAAGGGGGTACCTGCATATACCTCCAGGAAGAATGCTGTTTCTGTGTTAATGAATCTGG 1596

Qy        1236 AATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCG---AATACAATGTAGAGCAGAGGA 1292
                |||||   |     ||||       |||  ||  |      | |   | ||    | |
Db        1597 CATTGTTCATATCGCAGTTCGTAGGCTTCATGACAGGGCTGCAGAGCTTTGACATCAAGT 1656

Qy        1293 GCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTT 1352
                  | |   |      |   |      |||      |||||| |||||||||  ||||||
Db        1657 CGCTGACTCCTGGTGGCAAGGATCATCCCTTCTAAGATGGATACCCTGGGTTGCCCCCTT 1716

Qy        1353 CTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCT 1412
               | ||||||| ||       ||  |  |||||||  |  |||| || || || ||||||||
Db        1717 CCTAGGACCCCTGATCTTCCTCTTCCTGTTACTAATGATTGGGCCATGCATATTTAACCT 1776

RESULT 21
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 80595
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

  Query Match              6.1%;  Score 90.8;  DB 3;  Length 80595;
  Best Local Similarity   57.5%;  Pred. No. 8.4e-18;
  Matches  206;  Conservative    0;  Mismatches  142;  Indels   10;  Gaps    2;

Qy        1124 TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGG 1183
               | ||||||||    || ||||| ||  |||   ||||||  |     |  ||| || |
Db       58859 TGGCAGCAGTGACTCTCCAAAACCGCTGAGGCCTAGACTCTCATGTTGAGAAAGGAAG-- 58916

Qy        1184 GAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCA 1243
                 |        || ||||| | |||  | |||| ||||  ||| || ||  | || ||
Db       58917 --ATTCTGCACTTCTTAGGGGTAGAGTGTTGTTTTTATACTAACCAGTCAGGGATAGTAT 58974

Qy        1244 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACA 1303
                 || |      |     | |  ||             |    | |    |||  ||||
Db       58975 GAGATACCACCCAGTGTTTACAGGAAAAGGCTTCTGAAATCAGACAATGCCTTTCAAACT 59034

Qy        1304 CCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTC 1363
               |       |    |  |||| |    |||  | | ||| ||||||||||  |||| |||
Db       59035 C------TTATACCAACCTCTGGAGTTGGGCGACATGGCTTCTCCCCTTTCTAGGTCCTG 59088

Qy        1364 TAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGT 1423
               |  ||||  |  |  |  || ||   ||||| |||||||| ||||||||| | || || |
Db       59089 TGACAGCCATCTTGCTAATAGTCGCATTTGGGCCCTGTATTTTTAACCTCTTGGTCAAAT 59148

Qy        1424 TTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               |||| || || || || ||  | |||||||||||| ||||||||||| |||||||
Db       59149 TTGTTTCCTCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCA 59206

RESULT 22
US-09-120-653D-6
; Sequence 6, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
;  APPLICANT: YOON, JI-WON
;  APPLICANT:  JUN, HEE-SOOK
;  APPLICANT:  PARK, HAE-JOON
;  APPLICANT:  AHN, JONG-SEONG
;  APPLICANT:  HA, YOUNG-JU
;  APPLICANT:  CHUNG, SOO-IL
;  TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
;  FILE REFERENCE: 98-338
;  CURRENT APPLICATION NUMBER: US/09/120,653D
;  CURRENT FILING DATE:  1998-07-22
;  PRIOR APPLICATION NUMBER: KR 98-10108
;  PRIOR FILING DATE: 1998-03-24
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 6
;   LENGTH: 464
;   TYPE: DNA
;   ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-6

  Query Match              6.0%;  Score 89;  DB 3;  Length 464;
  Best Local Similarity   55.5%;  Pred. No. 1.3e-18;
  Matches  218;  Conservative    0;  Mismatches  165;  Indels   10;  Gaps    2;

Qy        1088 ACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATC 1147
               | ||  |  | || || |||    || |  || |  | ||||||||    |  ||||| |
Db          38 AATCTGTCCTTACTTTACAATCCAAAATAGACACTTTGGCAGCAGTGACTCGCCAAAACC 97

Qy        1148 GAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAG 1207
               |  |||   ||||  | || || ||  ||  | | |||  |||    || ||||| ||||
Db          98 GCTGAGGCATAGATGTCCTCACTGCTGAAGAAAGAGGACTCTGCACTTTCTTAGGGGAAG 157

Qy        1208 AACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAG 1267
               |  | |||| | |   ||| ||  |  | || |     ||          |  |  |
Db         158 AGTGTTGTTTTGACACTAACCAGGCACGGATAGCATGAGAT---------GGCACCCAGC 208

Qy        1268 ATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCC 1327
                |  |    ||      || ||   |  ||         |    |    |  |||  |
Db         209 GTTTACAGAAAAAGGCTGCTGAAATGAGTCGCCTTTTAAATTCTTATACCAACCTGTGGT 268

Qy        1328 AATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCC 1387
                  ||    | ||| ||||| ||||  |||| ||  | ||| |  |  |  ||||||||
Db         269 GTGGGGCAACATGGCTTCTCACCTTTCTAGGTCCCGTGGCAACCGTCTTGGTGTTACTCG 328

Qy        1388 TCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTG 1447
                |||||| ||| | || |||||| | ||||| || ||||| |  || ||| | || ||
Db      329 CCTTTGGGCCCAGCATTTTTAACGTTCTTGTCAAATTTGT-TTGTCTAGATTCGAGGCCT 387

Qy     1448 TAAAGCTACAGATGGTCTTACAAATGGAACCCC 1480
            | |||||||||||||||||||| ||  ||||||
Db      388 TCAAGCTACAGATGGTCTTACATATCAAACCCC 420

RESULT 23
US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
;  APPLICANT: YOON, JI-WON
;  APPLICANT:  JUN, HEE-SOOK
;  APPLICANT:  PARK, HAE-JOON
;  APPLICANT:  AHN, JONG-SEONG
;  APPLICANT:  HA, YOUNG-JU
;  APPLICANT:  CHUNG, SOO-IL
;  TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
;  FILE REFERENCE: 98-338
;  CURRENT APPLICATION NUMBER: US/09/120,653D
;  CURRENT FILING DATE:  1998-07-22
;  PRIOR APPLICATION NUMBER: KR 98-10108
;  PRIOR FILING DATE: 1998-03-24
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 1
;   LENGTH: 3910
;   TYPE: DNA
;   ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

  Query Match             6.0%;  Score 89;  DB 3;  Length 3910;
  Best Local Similarity   55.5%;  Pred. No. 4.9e-18;
  Matches  218;  Conservative    0;  Mismatches  165;  Indels   10;  Gaps    2;

Qy     1088 ACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATC 1147
            | ||  |  | || || |||    || |  || |  | ||||||||    |  ||||| |
Db     2895 AATCTGTCCTTACTTTACAATCCAAAATAGACACTTTGGCAGCAGTGACTCGCCAAAACC 2954

Qy     1148 GAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAG 1207
            |  |||   ||||  | || || ||  ||  | | |||  |||    || ||||| ||||
Db     2955 GCTGAGGCATAGATGTCCTCACTGCTGAAGAAAGAGGACTCTGCACTTTCTTAGGGGAAG 3014

Qy     1208 AACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAG 1267
            |  | |||| | |   ||| ||  |  | || |     ||          |  |  |
Db     3015 AGTGTTGTTTTGACACTAACCAGGCACGGATAGCATGAGAT---------GGCACCCAGC 3065

Qy     1268 ATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCC 1327
             |  |    ||      || ||   |  ||         |    |    |  |||  |
Db     3066 GTTTACAGAAAAAGGCTGCTGAAATGAGTCGCCTTTTAAATTCTTATACCAACCTGTGGT 3125

Qy     1328 AATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCC 1387
               ||    | ||| ||||| |||| ||||| ||  | ||| |  |  |  ||||||||
Db     3126 GTGGGGCAACATGGCTTCTCACCTTTCTAGGTCCCGTGGCAACCGTCTTGGTGTTACTCG 3185

Qy     1388 TCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTG 1447
             |||||| ||| | || |||||| | |||||| || ||||| |  || ||| | || ||
Db     3186 CCTTTGGGCCCAGCATTTTTAACGTTCTTGTCAAATTTGT-TTGTCTAGATTCGAGGCCT 3244

Qy     1448 TAAAGCTACAGATGGTCTTACAAATGGAACCCC 1480
            | |||||||||||||||||||| ||  ||||||
Db     3245 TCAAGCTACAGATGGTCTTACATATCAAACCCC 3277

RESULT 24
US-09-315-127-7
; Sequence 7, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
;  APPLICANT: The University of Tennessee, c/o Richard Cox
;  TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
;  TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
;  FILE REFERENCE: 44137-5023, U. of Tennessee
;  CURRENT APPLICATION NUMBER: US/09/315,127
;  CURRENT FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 2002
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SEQ. ID NO.
;   OTHER INFORMATION: 5, retroviral vector
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (65)..(1999)
US-09-315-127-7

  Query Match             4.7%;  Score 70.2;  DB 4;  Length 2002;
  Best Local Similarity   48.7%;  Pred. No. 4.2e-12;
  Matches  235;  Conservative    0;  Mismatches  233;  Indels   15;  Gaps    1;

Qy     1014 CAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACAT 1073
            || |  |  || |   || |||   |   |  | ||      |  ||  |    ||| |
Db     1492 CACTGCCCTAATCAAAACCCAGCAGTTTGAGCAGCTTCACGCCGCTATCCAGACAGACCT 1551

Qy     1074 GGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGT 1133
              |  | |||    | ||  |   || | |  || |   ||| | |||  |  | | |||
Db     1552 CAACGAAGTCGAAAAATCAATTACCAACCTAGAAAAGTCACTGACCTCGTTGTCTGAAGT 1611

Qy     1134 AGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTT 1193
            |||||| || || |||||||   |||| |||||   |   |||   || ||   |||
Db     1612 AGTCCTACAGAACCGAAGAGGCCTAGATTTGCTCTTCCTAAAAGAGGGAGGTCTCTGCGC 1671

Qy     1194 ATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGT 1253
            |    ||  ||||||| | |||| |||||   | ||  |  || | || |  || |   |
Db     1672 AGCCCTAAAAGAAGAATGTTGTTTTTATGCAGACCACACGGGACTAGTGAGAGACAGCAT 1731

Qy     1254 TAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTG 1313
                 || |  | ||  |  |  |    ||   |  | |      || |     | || | ||
Db     1732 GGCCAAACTAAGGGAAAGGCTTAATCAGAGACAAAAACTATTTGAGTCAGGCCAAGGTTG 1791

Qy     1314 GGGCCTCCT---------------CAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGG 1358
            |  |                        |  |   |  |   ||   |   |   || || | ||
Db     1792 GTTCGAAGGGCAGTTTAATAGATCCCCCTGGTTTACCACCTTAATCTCCACCATCATGGG 1851

Qy     1359 ACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGT 1418
            |||||||  ||  ||  || || |  | ||||||||||||| ||  | || |   | ||
Db     1852 ACCTCTAATAGTACTCTTACTGATCTTACTCTTTGGACCCTGCATTCTCAATCGATTGGT 1911

Qy     1419 TAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACC 1478
              | |||||      ||| ||   ||  ||  ||   | | |  |    |||    | |
Db     1912 CCAATTTGTTAAAGACAGGATCTCAGTGGTCCAGGCTCTGGTTTTGACTCAACAATATCA 1971

Qy     1479 CCA 1481
            |||
Db     1972 CCA 1974

RESULT 25
US-09-513-999C-24228
; Sequence 24228, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 24228
;   LENGTH: 149
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 127
;   OTHER INFORMATION: r=a or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 129
;   OTHER INFORMATION: n=a, g, c or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 135
;   OTHER INFORMATION: m=a or c
US-09-513-999C-24228

  Query Match              4.7%;  Score 70;  DB 4;  Length 149;
  Best Local Similarity   88.4%;  Pred. No. 9.6e-13;
  Matches   76;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1396 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1455
              | ||||||||||||||||||||||||||||||||||||||||||  ||||| |||| |||
Db          5 CTCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTATAAAACTA 64

Qy       1456 CAGATGGTCTTACAAATGGAACCCCA 1481
              || |||||  | | |||||||| |||||
Db         65 CAAATGGTTCTCCAAATGGAGCCCCA 90


RESULT 26
US-09-011-745-9
; Sequence 9, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 3925
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3910)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3911)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3912)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (3913)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-9

  Query Match              4.7%;  Score 70;  DB 3;  Length 3925;
  Best Local Similarity   49.3%;  Pred. No. 7.5e-12;
  Matches  224;  Conservative    0;  Mismatches  215;  Indels   15;  Gaps    1;

Qy       1014 CAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACAT 1073
              || |  |  || |   || ||   |   |  | ||       |  ||  ||    ||| |
Db       2948 CACTGCCCTAATCAAAACCCAGCAGTTTGAGCAGCTTCACGCCGCTATCCAGACAGACCT 3007

Qy       1074 GGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGT 1133
                |  | |||    | ||  ||   || | |  || |   ||| | |||  |  | | |||
Db       3008 CAACGAAGTCGAAAAATCAATTACCAACCTAGAAAAGTCACTGACCTCGTTGTCTGAAGT 3067

Qy       1134 AGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTT 1193
              |||||| || || |||||||   |||| ||||||   |   |||   || ||   |||   
Db       3068 AGTCCTACAGAACCGAAGAGGCCTAGATTTGCTCTTCCTAAAAGAGGGAGGTCTCTGCGC 3127

Qy       1194 ATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGT 1253
              |    ||  ||||||| | |||| |||||   | ||  |  || | || |  || |   |
Db       3128 AGCCCTAAAAGAAGAATGTTGTTTTTATGCAGACCACACGGGACTAGTGAGAGACAGCAT 3187

Qy       1254 TAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTG 1313
                   || |  | ||  |  |  |   |||  | |     || |      | || | ||
Db       3188 GGCCAAACTAAGGGAAAGGCTTAATCAGAGACAAAAACTATTTGAGTCAGGCCAAGGTTG 3247

Qy       1314 GGGCCTCCT---------------CAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGG 1358
              |  |                        |  |   |  |   |||   |   | || || | ||
Db       3248 GTTCGAAGGGCAGTTTAATAGATCCCCCTGGTTTACCACCTTAATCTCCACCATCATGGG 3307

Qy       1359 ACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGT 1418
              |||||||  ||  ||  || || |  | |||||||||||||| ||  | || | |   | ||
Db       3308 ACCTCTAATAGTACTCTTACTGATCTTACTCTTTGGACCCTGCATTCTCAATCGATTAGT 3367

Qy       1419 TAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAG 1452
              | | |||||      ||| ||   ||  ||  ||
Db       3368 TCAATTTGTTAAAGACAGGATCTCAGTAGTCCAG 3401


RESULT 27
US-08-979-847B-102
; Sequence 102, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLA
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 102:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 635 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-08-979-847B-102

  Query Match             4.6%;  Score 68.4;  DB 4;  Length 635;
  Best Local Similarity   87.2%;  Pred. No. 7.9e-12;
  Matches   75;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1396 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 1455
              ||||||||||||||||||||||||||||||||||||||||||||  || ||||||| |||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy       1456 CAGATGGTCTTACAAATGGAACCCCA 1481
              || || ||  | |||||||| | |||
Db         61 CAAATTGTTCTTCAAATGGAGCACCA 86


RESULT 28
US-09-011-745-5
; Sequence 5, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 6028
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3774)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3775)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3776)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3777)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3962)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3963)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3964)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3965)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-5

  Query Match             4.6%;  Score 68.4;  DB 3;  Length 6028;
  Best Local Similarity   49.5%;  Pred. No. 3.2e-11;
  Matches  217;  Conservative    0;  Mismatches  206;  Indels   15;  Gaps    1;

Qy       1030 ACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGAC 1089
              || |||   |   |  | ||   |   |  ||  |    ||| |  |  | |||    | 
Db       2828 ACCCAGCAGTTTGAGCAGCTTCATGCCGCTATCCAGACAGACCTCAACGAAGTCGAAAAG 2887

Qy       1090 TCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGA 1149
              ||  |   || | |  || |   ||| | |||  |  | | ||||||||| || || || 
Db       2888 TCAATTACCAACCTAGAAAAGTCACTGACCTCGTTGTCTGAAGTAGTCCTACAGAACCGC 2947

Qy       1150 AGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAA 1209
              ||||   |||| ||||||  |   ||    || ||   |||   |    ||  |||||||
Db       2948 AGAGGCCTAGATTTGCTATTCCTAAAGGAGGGAGGTCTCTGCGCAGCCCTAAAAGAAGAA 3007

Qy       1210 CGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGAT 1269
               | |||| |||||   | ||  |  |  | || |  || |   |     || |  |||| 
Db       3008 TGTTGTTTTTATGCAGACCACACGGGGCTAGTGAGAGACAGCATGGCCAAATTAAGAGAA 3067

Qy       1270 CGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCT------- 1322
               |  |  |    |||  | |     || | |    | || | |||  |          
Db       3068 AGGCTTAATCAGAGACAAAAACTATTTGAGACAGGCCAAGGATGGTTCGAAGGGCTGTTT 3127

Qy       1323 --------CAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTA 1374
                      |  |   |  |   |||   |   | || || | ||||||||||  ||  || 
Db       3128 AATAGATCCCCCTGGTTTACCACCTTAATCTCCACCATCATGGGACCTCTAATAGTACTC 3187

Qy       1375 ATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCC 1434
               || || |  | ||||||||||| || ||  | || |   | ||| | ||||||      |
Db       3188 TTACTGATCTTACTCTTTGGACCTTGCATTCTCAATCGATTAGTTCAATTTGTTAAAGAC 3247

Qy       1435 AGAATTGAAGCTGTAAAG 1452
              || ||   ||  ||  ||
Db       3248 AGGATCTCAGTAGTCCAG 3265


RESULT 29
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1

  Query Match             4.6%;  Score 68.4;  DB 3;  Length 8060;
  Best Local Similarity   53.3%;  Pred. No. 3.9e-11;
  Matches  144;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy       1045 AAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTG 1104
              || |||  || |    |||  |  ||  |  ||    |     | ||  |    | | ||
Db       1523 AACCTACATCGAATTGTAACAGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCTG 1582

Qy       1105 CAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTG 1164
               | ||    || | |||| || | | ||||||||| || ||| ||||||  ||||| ||
Db       1583 GAGGAATCCCTAACCTCCTTATCTGAAGTAGTCCTACAGAATAGAAGAGGGTTAGATTTA 1642

Qy       1165 CTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTT 1224
               ||      |||  ||| |||   ||| ||   ||    || ||| |||||| ||||||
Db       1643 TTATTTCTAAAAGAAGGAGGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTG 1702

Qy       1225 AATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGA 1284
               |||| ||  |     |||  ||     | ||  || || ||||  |  |  |     |
Db       1703 GATCATTCAGGGGCCATCAGAGACTCCATGAACAAACTTAGAGAAAGGTTGGAGAAGCGT 1762

Qy       1285 GCAGAGGAGCTTCAAAACACCGAACGCTGG 1314
                |  |||     |||  ||  || | |||
Db       1763 CGAAGGGAAAAGGAAACTACTCAAGGGTGG 1792


RESULT 30
US-09-661-858-1
; Sequence 1, Application US/09661858
; Patent No. 6699663
;   GENERAL INFORMATION:
;        APPLICANT: Jay A. Fishman
;        TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;                            AND METHODS OF USE
;        NUMBER OF SEQUENCES: 74
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: LAHIVE & COCKFIELD, LLP
;             STREET: 60 State Street
;             CITY: Boston
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02109-1875
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
```

```
;             APPLICATION NUMBER: US/09/661,858
;             FILING DATE: 14-Sep-2000
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/766,528
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Louis Myers
;             REGISTRATION NUMBER: 35,965
;             REFERENCE/DOCKET NUMBER: MGP-038CP
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617)227-7400
;             TELEFAX: (617)227-5941
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 8060 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-661-858-1

  Query Match             4.6%;  Score 68.4;  DB 4;  Length 8060;
  Best Local Similarity   53.3%;  Pred. No. 3.9e-11;
  Matches  144;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy       1045 AAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTG 1104
              || |||  || |    |||  |  ||  |  ||    |     | ||  |    | | ||
Db       1523 AACCTACATCGAATTGTAACAGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCTG 1582

Qy       1105 CAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTG 1164
               | ||    || | |||| || | | ||||||||| || ||| ||||||  ||||| ||
Db       1583 GAGGAATCCCTAACCTCCTTATCTGAAGTAGTCCTACAGAATAGAAGAGGGTTAGATTTA 1642

Qy       1165 CTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTT 1224
               ||      |||  ||| |||   ||| ||   ||    || ||| |||||| ||||||
Db       1643 TTATTTCTAAAAGAAGGAGGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTG 1702

Qy       1225 AATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGA 1284
               |||| ||  |     |||  ||     | ||  || || ||||  |  |  |     |
Db       1703 GATCATTCAGGGGCCATCAGAGACTCCATGAACAAACTTAGAGAAAGGTTGGAGAAGCGT 1762

Qy       1285 GCAGAGGAGCTTCAAAACACCGAACGCTGG 1314
                |  |||     |||  ||  || | |||
Db       1763 CGAAGGGAAAAGGAAACTACTCAAGGGTGG 1792


Search completed: January 22, 2005, 19:36:59
Job time : 141.138 secs
```

```
                         GenCore version 5.1.6
              Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 12:12:40 ; Search time 804.34 Seconds
                                            (without alignments)
                                            10579.682 Million cell updates/sec

Title:          US-09-319-156B-9
Perfect score:  1481
Sequence:       1 atggccctcccttatcatac..........gtcttacaaatggaacccca 1481

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters:      8600550

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                 1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                 2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                 3:  /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                 4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                 5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                 6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                 7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                 8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                 9:  /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                 10:  /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
                 11:  /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
                 12:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                 13:  /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
                 14:  /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
                 15:  /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
                 16:  /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
                 17:  /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
                 18:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                 19:  /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
                 20:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                 21:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                           Description
---------------------------------------------------------------------------
    1    1481   100.0   1481  8  US-08-979-847-105          Sequence 105, App
    2    1481   100.0   1481 15   US-10-114-104-105          Sequence 105, App
    3    1481   100.0   2030 16   US-10-637-565-18           Sequence 18, Appl
    4  1389.8    93.8   2074 16   US-10-416-642-4            Sequence 4, Appli
    5  1335.4    90.2   2930  9  US-09-902-535-1            Sequence 1, Appli
    6  1335.4    90.2   2946 13   US-10-114-893-134          Sequence 134, App
    7  1335.4    90.2   2946 15   US-10-016-249-3            Sequence 3, Appli
    8  1333.8    90.1   1617 16   US-10-133-036-3            Sequence 3, Appli
    9  1333.8    90.1   8523 10   US-09-854-867-21           Sequence 21, Appl
   10  1333.8    90.1  56093 10   US-09-873-367C-81          Sequence 81, Appl
   11  1324.2    89.4   7582 16   US-10-632-793-30           Sequence 30, Appl
   12  1308.2    88.3   2782 16   US-10-133-036-1            Sequence 1, Appli
```

```
      13  1306.6   88.2   2782  16  US-10-632-793-26          Sequence 26, Appl
 c    14  1277.4   86.3 161334  13  US-10-087-192-730         Sequence 730, App
      15  1131.6   76.4   1894  9  US-09-864-761-4444        Sequence 4444, Ap
      16  1116.4   75.4   1948  16  US-10-632-793-24          Sequence 24, Appl
      17   716.6   48.4    792  9  US-09-864-761-21192       Sequence 21192, A
      18   591.8   40.0   1684  16  US-10-363-616-228         Sequence 228, App
      19   582.8   39.4    591  16  US-10-632-793-19          Sequence 19, Appl
      20   516.6   34.9   1136  16  US-10-632-793-25          Sequence 25, Appl
      21   513.6   34.7   2006  16  US-10-632-793-23          Sequence 23, Appl
 c    22   377.8   25.5    570  13  US-10-027-632-322491      Sequence 322491,
 c    23   377.8   25.5    570  13  US-10-027-632-322492      Sequence 322492,
 c    24   377.8   25.5    570  15  US-10-027-632-322491      Sequence 322491,
 c    25   377.8   25.5    570  15  US-10-027-632-322492      Sequence 322492,
 c    26     364   24.6    521  15  US-10-029-386-4312        Sequence 4312, Ap
      27     353   23.8    551  13  US-10-027-632-322574      Sequence 322574,
      28     353   23.8    551  13  US-10-027-632-322575      Sequence 322575,
      29     353   23.8    551  15  US-10-027-632-322574      Sequence 322574,
      30     353   23.8    551  15  US-10-027-632-322575      Sequence 322575,
 c    31   318.6   21.5 822900  15  US-10-292-798-1393        Sequence 1393, Ap
      32     302   20.4   1329  8  US-08-979-847-108         Sequence 108, App
      33     302   20.4   1329  15  US-10-114-104-108         Sequence 108, App
      34     296   20.0    426  9  US-09-864-761-14030       Sequence 14030, A
 c    35   229.8   15.5    625  13  US-10-027-632-316412      Sequence 316412,
 c    36   229.8   15.5    625  15  US-10-027-632-316412      Sequence 316412,
 c    37   229.6   15.5    619  13  US-10-027-632-86541       Sequence 86541, A
 c    38   229.6   15.5    619  15  US-10-027-632-86541       Sequence 86541, A
 c    39   219.4   14.8    559  9  US-09-864-761-7501        Sequence 7501, Ap
      40     213   14.4    494  15  US-10-029-386-20259       Sequence 20259, A
      41   194.8   13.2    520  9  US-09-864-761-7310        Sequence 7310, Ap
      42   189.8   12.8    740  13  US-10-027-632-150080      Sequence 150080,
      43   189.8   12.8    740  15  US-10-027-632-150080      Sequence 150080,
 c    44   173.8   11.7    527  13  US-10-027-632-85218       Sequence 85218, A
 c    45   173.8   11.7    527  15  US-10-027-632-85218       Sequence 85218, A

                                  ALIGNMENTS


RESULT 1
US-08-979-847-105
; Sequence 105, Application US/08979847
; Publication No. US20030039664A1
;  GENERAL INFORMATION:
;    APPLICANT:  PERRON, HERVE
;    APPLICANT:  BESEME, FREDERIC
;    APPLICANT:  BEDIN, FREDERIC
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  KOMURIAN-PRADEL, FLORENCE
;    APPLICANT:  JOLIVET-REYNAUD, COLETTE
;    APPLICANT:  MANDRAND, BERNARD
;    APPLICANT:  GARSON, JEREMY
;    APPLICANT:  TUKE, PHILIP
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  210
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/979,847
;      FILING DATE:  26-NOV-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  BERRIDGE, WILLIAM P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 39046A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  105:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1481 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-979-847-105

  Query Match             100.0%;  Score 1481;  DB 8;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 2
US-10-114-104-105
; Sequence 105, Application US/10114104
; Publication No. US20030198647A1
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLA
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
```

```
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/114,104
;             FILING DATE: 03-Apr-2002
;             CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/979,847
;             FILING DATE: 26-NOV-1997
;         ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;     INFORMATION FOR SEQ ID NO: 105:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 1481 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 105:
US-10-114-104-105

  Query Match             100.0%;  Score 1481;  DB 15;  Length 1481;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 3
US-10-637-565-18
; Sequence 18, Application US/10637565
; Publication No. US20040043381A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  PERRON, Herve
;  APPLICANT:  KOMURIAN-PRADEL, Florence
;  TITLE OF INVENTION: THE LTR REGION OF MSRV-1 AND THE PROTEINS IT ENCODES, AND PROBES
;  TITLE OF INVENTION:  METHODS FOR DETECTING MSRV-1 RETROVIRUS
;  FILE REFERENCE: 110257
;  CURRENT APPLICATION NUMBER: US/10/637,565
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: US/09/890,340
;  PRIOR FILING DATE: 2001-11-27
;  PRIOR APPLICATION NUMBER: PCT/IB00/00159
;  PRIOR FILING DATE: 2000-02-15
;  PRIOR APPLICATION NUMBER: EP 99420041.8
;  PRIOR FILING DATE: 1999-02-15
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 2030
;   TYPE: DNA
;   ORGANISM: MSRV-1 retrovirus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1626)
US-10-637-565-18

  Query Match             100.0%;  Score 1481;  DB 16;  Length 2030;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1481;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 4
US-10-416-642-4
; Sequence 4, Application US/10416642
; Publication No. US20040043452A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  RAMKUMAR, Jayalaxmi
;  APPLICANT:  ARVIZU, Chandra
;  TITLE OF INVENTION: EMBRYOGENESIS ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-0842 PCT
;  CURRENT APPLICATION NUMBER: US/10/416,642
;  CURRENT FILING DATE:  2003-05-13
;  PRIOR APPLICATION NUMBER: 60/249,407
;  PRIOR FILING DATE: 2000-11-15
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PERL Program
; SEQ ID NO 4
;   LENGTH: 2074
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20040043452A1 7477736CB1
;   FEATURE:
```

```
;   NAME/KEY: unsure
;   LOCATION: 1995
;   OTHER INFORMATION: a, t, c, g, or other
US-10-416-642-4

  Query Match             93.8%;  Score 1389.8;  DB 16;  Length 2074;
  Best Local Similarity   96.2%;  Pred. No. 0;
  Matches 1424;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              |||||||||||||||  || |||||||||||||||||||||| ||| ||||||  |||||||
Db         64 ATGGCCCTCCCTTATTGTATTTTTCTCTTTACTGTTCTCTCACCACCTTTCAGTCTCACT 123

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              |||||| ||||||||  |||||  |||||||||||||||||||||| ||||||||||| |||
Db        124 GCACCCTCTCCATGCCACTGTAGGACCAGTAGCTCCCCTTACCAAGCGTTTCTATGGAGA 183

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | ||||| |||  || |||||||||||||||| ||||||||||||||||||||||||| ||
Db        184 ATGCGGCGTCCCAGACATATTGATGCCCCATCGTATAGGAGTTTATCTAAGGGAAACCCC 243

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||| |||||||||||||||||||| |||||||||||||||||||||||||| |||
Db        244 GCCTTCACCGCCCACACCCATATGCCCCACAACTGCTATAACTCTGCCACTCTTTGTATG 303

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 363

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||| |||||||||||||||||||||||||| |||||||||||||||||||| ||
Db        364 GGAGCCACTATCTGTTGGACTTACTTCACCCATACCGGTATGTCTGATGGGGGTGGAGTT 423

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| ||||||||||||||||||| |||||||||| ||||||||||||| ||||||| ||||
Db        424 CAAGATCAGGCAAGAGAAAAACATGTAAAGGAAGTAATCTCCCAACTCACCCGGGTACAT 483

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        484 AGCACCCCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 543

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              ||||||| ||||||| ||||||||||||||||||||||||| ||||||||||||||||| |||
Db        544 CATACTCACCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 603

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        604 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAGTC 663

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| |||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Db        664 CCTGTACCTGAACAATGGAATAACTTCAGCACAGAAATAAACACCACTTCCATTTTAGTA 723

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        724 GGACCTCTTGTTTCCAATCTGGAAATAATCCATACCTCAAACCTCACCTGTGTAAAATTT 783

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||||||  ||||||| |||||| |||||||| ||||||||| ||||||||||| |
Db        784 AGCAATACTATAGCAACAACCAACTCCCAGTGCATCAGATGGGTAACTCCTCCCACACAA 843

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 903

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 963

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 1023
```

```
Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1083

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1143

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1203

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGAACCTGTTTATTTTTA 1263

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1264 GGAGAAGAATGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1323

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1324 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACACTGGGGCCTC 1383

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1384 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1443

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1444 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1503

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||||||||| |||||  | |||||||| |||||
Db       1504 GAAGCTGTAAAGCTACAAATGGTTCTTCAAATGGAGCCCCA 1544

RESULT 5
US-09-902-535-1
; Sequence 1, Application US/09902535
; Patent No. US20020102530A1
; GENERAL INFORMATION:
;  APPLICANT: Keith, Jr., James C.
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  Mi, Sha
;  TITLE OF INVENTION: Methods and compositions for diagnosing
;  TITLE OF INVENTION:  and treating preeclampsia and gestational trophoblast
;  TITLE OF INVENTION:  disorders
;  FILE REFERENCE: GIN-6006B4
;  CURRENT APPLICATION NUMBER: US/09/902,535
;  CURRENT FILING DATE:  2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,657
;  PRIOR FILING DATE: 2000-07-06
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2930
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (930)...(2546)
US-09-902-535-1

  Query Match             90.2%;  Score 1335.4;  DB 9;  Length 2930;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||||| |||||| ||||| ||||||||
Db        930 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 989

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
```

```
             ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       990 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1049

Qy       121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
             | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db      1050 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1109

Qy       181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
             ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      1110 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1169

Qy       241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
             |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db      1170 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1229

Qy       301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
             |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db      1230 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1289

Qy       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
             |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db      1290 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1349

Qy       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      1350 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1409

Qy       481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
             |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db      1410 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1469

Qy       541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
             ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db      1470 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1529

Qy       601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
             ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1530 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1589

Qy       661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1590 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1649

Qy       721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
             |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db      1650 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1709

Qy       781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      1710 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1769

Qy       841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1770 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1829

Qy       901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
             ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db      1830 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1889

Qy       961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
             ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db      1890 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1949

Qy      1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
             ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db      1950 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2009

Qy      1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
             ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2010 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2069

Qy      1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
             ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
```

```
Db      2070 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2129

Qy      1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
             || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db      2130 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2189

Qy      1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
             |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db      2190 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2249

Qy      1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
             ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db      2250 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2309

Qy      1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db      2310 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2369

Qy      1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
             ||||||||||| ||||| ||||      | ||| |  || |
Db      2370 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2410
```

```
RESULT 6
US-10-114-893-134
; Sequence 134, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT: McCoy, John M.
;  APPLICANT: LaVallie, Edward R.
;  APPLICANT: Collins-Racie, Lisa A.
;  APPLICANT: Evans, Cheryl
;  APPLICANT: Merberg, David
;  APPLICANT: Treacy, Maurice
;  APPLICANT: Bowman, Michael R.
;  APPLICANT: Spaulding, Vikki
;  APPLICANT: Carlin-Duckett, McKeough
;  APPLICANT: Kelleher, Kerry S.
;  APPLICANT: Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: GI 6000-10A
;  CURRENT APPLICATION NUMBER: US/10/114,893
;  CURRENT FILING DATE: 2002-04-02
;  EARLIER APPLICATION NUMBER: 09/413,232
;  EARLIER FILING DATE: 1999-10-06
;  NUMBER OF SEQ ID NOS: 321
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 134
;   LENGTH: 2946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-114-893-134

  Query Match             90.2%;  Score 1335.4;  DB 13;  Length 2946;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy         1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
             ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db       928 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 987

Qy        61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
             ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       988 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1047

Qy       121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
             | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db      1048 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1107

Qy       181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
             ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      1108 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1167
```

```
Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        1168 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1227

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        1228 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1287

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        1288 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1347

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        1348 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1407

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        1408 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1467

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        1468 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1527

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1528 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1587

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1588 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1647

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db        1648 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1707

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        1708 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1767

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1768 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1827

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||||| ||||| || ||||||||| |||||||||||||||||||
Db        1828 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1887

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        1888 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1947

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1948 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2007

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2008 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2067

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        2068 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2127

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db        2128 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2187

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db        2188 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2247

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        2248 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2307

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        2308 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2367

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||| ||||| ||||      | ||| |  || |
Db        2368 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2408

RESULT 7
US-10-016-249-3
; Sequence 3, Application US/10016249
; Publication No. US20030100053A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Mi, Sha
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6006B.AJ172A
;  CURRENT APPLICATION NUMBER: US/10/016,249
;  CURRENT FILING DATE:  2001-10-30
;  PRIOR APPLICATION NUMBER: US/09/175,928
;  PRIOR FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-016-249-3

  Query Match             90.2%;  Score 1335.4;  DB 15;  Length 2946;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 1390;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db         928 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 987

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         988 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1047

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        1048 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1107

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        1108 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1167

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        1168 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1227

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        1228 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1287

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        1288 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1347
```

```
Qy     421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
            ||||| ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    1348 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1407

Qy     481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
           |||||||||||||| ||||||||||||||||||||||||| |||||||||||||||| |||
Db    1408 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1467

Qy     541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
           ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db    1468 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1527

Qy     601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1528 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1587

Qy     661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1588 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1647

Qy     721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
           |||||||||| | ||||||||| ||||||||||||||||||||||||| ||||||||| |
Db    1648 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1707

Qy     781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    1708 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1767

Qy     841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1768 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1827

Qy     901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
           ||||||||||||||||||||| ||||| || ||||||||| ||||||||||||||||||||
Db    1828 ACTGAACAAGATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1887

Qy     961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
           ||||||||||||||  ||||||||||||||||||   ||||||||||||||||| |||||
Db    1888 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1947

Qy    1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
           ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db    1948 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 2007

Qy    1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
           ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2008 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 2067

Qy    1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
           ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db    2068 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 2127

Qy    1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
           || |||||| |||||||||||||||||||||||| |||| ||||||||||||||||||||
Db    2128 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2187

Qy    1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
           |||||||||||||||||| ||||||||||||||||||| ||||||| | || ||||||||||
Db    2188 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2247

Qy    1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
           ||||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db    2248 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2307

Qy    1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
            ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db    2308 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2367

Qy    1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
           ||||||||||| ||||| ||||      | ||| |  || |
Db    2368 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2408
```

```
RESULT 8
US-10-133-036-3
; Sequence 3, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
;  APPLICANT: Conrad, Bernard
;  APPLICANT:  Mach, Bernard
;  TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
;  FILE REFERENCE: 23135-507
;  CURRENT APPLICATION NUMBER: US/10/133,036
;  CURRENT FILING DATE:  2002-04-26
;  PRIOR APPLICATION NUMBER: PCT/EP00/10659
;  PRIOR FILING DATE: 2000-10-30
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1617
;   TYPE: DNA
;   ORGANISM: Human endogenous retrovirus
US-10-133-036-3

  Query Match             90.1%;  Score 1333.8;  DB 16;  Length 1617;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy       1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
           ||||||||||||||||||| |||||||||||||||||||| ||||| |||||| ||||||||
Db       1 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 60

Qy      61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
           ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db      61 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 120

Qy     121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
           | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db     121 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 180

Qy     181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 240

Qy     241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     241 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300

Qy     301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
           |||| ||||||||||||||||||||||||||||| ||  |||||||||||||||||| ||
Db     301 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 360

Qy     361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
           |||| ||||||||||||||||||| ||||| |||| ||||||||||||| |||||| ||||
Db     361 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 420

Qy     421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
            ||||| ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     421 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 480

Qy     481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
           |||||||||||||| ||||||||||||||||||||||||| |||||||||||||||| |||
Db     481 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 540

Qy     541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
           ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db     541 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 600

Qy     601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660

Qy     661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
```

```
Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db        721 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 780

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 840

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db        901 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 960

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        961 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1020

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1080

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1200

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       1201 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1260

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       1261 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1320

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       1321 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1380

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       1381 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1440

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       1441 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 1481

RESULT 9
US-09-854-867-21
; Sequence 21, Application US/09854867
; Publication No. US20030224356A1
; GENERAL INFORMATION:
;  APPLICANT: JOAN, KNOLL H
;  APPLICANT:  ROGAN, PETER K
;  TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
;  FILE REFERENCE: 30307
;  CURRENT APPLICATION NUMBER: US/09/854,867
;  CURRENT FILING DATE:  2003-05-08
;  NUMBER OF SEQ ID NOS: 613
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
;   LENGTH: 8523
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(8523)
;   OTHER INFORMATION: herv17
US-09-854-867-21

  Query Match             90.1%;  Score 1333.8;  DB 10;  Length 8523;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db       6849 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 6908

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       6909 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 6968

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db       6969 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 7028

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       7029 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 7088

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       7089 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 7148

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       7149 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 7208

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       7209 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 7268

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       7269 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 7328

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       7329 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 7388

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       7389 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 7448

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7449 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 7508

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7509 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 7568

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       7569 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 7628

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       7629 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 7688

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7689 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 7748

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db       7749 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 7808

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       7809 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 7868

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       7869 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 7928

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7929 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 7988

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       7989 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 8048

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       8049 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 8108

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db       8109 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 8168

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       8169 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 8228

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db       8229 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 8288

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       8289 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 8329

RESULT 10
US-09-873-367C-81
; Sequence 81, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  APPLICANT:  Soppet, Daniel
;  APPLICANT:  Endress, Gregory
;  APPLICANT:  Augustus, Meena
;  APPLICANT:  Ebner, Reinhard
;  APPLICANT:  Carter, Kenneth
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
;  TITLE OF INVENTION:  Signature Gene Sets
;  FILE REFERENCE: 689290-64
;  CURRENT APPLICATION NUMBER: US/09/873,367C
;  CURRENT FILING DATE:  2003-04-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,891
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,842
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/244,867
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: U.S. 60/245,084
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 1067
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 81
;   LENGTH: 56093
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-873-367C-81

  Query Match            90.1%;  Score 1333.8;  DB 10;  Length 56093;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db      35879 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 35938

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db      35939 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 35998

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db      35999 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 36058

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      36059 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 36118

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db      36119 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 36178

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db      36179 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 36238

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db      36239 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 36298

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      36299 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 36358

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db      36359 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 36418

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db      36419 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 36478

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36479 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 36538

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36539 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 36598

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db      36599 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 36658

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      36659 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 36718

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36719 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 36778

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db      36779 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 36838

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db      36839 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 36898

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db      36899 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 36958

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36959 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 37018
```

```
Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db      37019 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 37078

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db      37079 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 37138

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db      37139 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 37198

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db      37199 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 37258

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db      37259 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 37318

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db      37319 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 37359


RESULT 11
US-10-632-793-30
; Sequence 30, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
;   LENGTH: 7582
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (307)..(307)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (355)..(355)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1309)..(1309)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1331)..(1331)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2213)..(2213)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2398)..(2398)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3787)..(3787)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4115)..(4115)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4261)..(4261)
;   OTHER INFORMATION: n = a or g or c or t/u
US-10-632-793-30

  Query Match             89.4%;  Score 1324.2;  DB 16;  Length 7582;
  Best Local Similarity   92.2%;  Pred. No. 0;
  Matches 1365;  Conservative   24;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||:|||||||||||||| ||||||||:||:||||:| |:|||| :|||| ||||||||
Db       5581 ATGGSCCTCCCTTATCATATTTTTCTCTKTASTGTTSTTTYACCCTSTTTCACTCTCACT 5640

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||:|||:|||||||||||| |||
Db       5641 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTYACCMAGAGTTTCTATGGAGA 5700

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||||| ||||||||| | |:|||||||| | ||
Db       5701 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTSTAAGGGAACCCCC 5760

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       5761 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 5820

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       5821 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 5880

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       5881 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 5940

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| |||||  |||| |||||||||||| |||:||| ||||
Db       5941 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCSGGGTACAT 6000

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| |||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       6001 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 6060

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       6061 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 6120

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| ||||||:||||||  ||||||||
Db       6121 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCARGCCATATGTTTCAATC 6180

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6181 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 6240

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       6241 GGACCTCTTGTTTCCAATSTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 6300

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
```

```
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       6301 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 6360

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       6361 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 6420

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              |||||||||||||||||||||||||||||||||||||||||||||||:|||:||||||||
Db       6421 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCYATGRCCATCTAC 6480

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db       6481 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 6540

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       6541 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 6600

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       6601 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 6660

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6661 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 6720

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              |:||||||||||||||||||||:|||||||||  |:||||||||||||||||||||||||
Db       6721 CRAAATCGAAGAGCTTTAGACTYGCTAACCGCTGARAGAGGGGGAACCTGTTTATTTTTA 6780

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| ||||||||||||||||||||||| |||| |||||||||||||||:|||||
Db       6781 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTRAAGAA 6840

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||:||||||||||||  |||:||||||:||||||| |||||| | || ||||||||||
Db       6841 ATTCSAGATCGAATACAACGTAKAGCAGARGAGCTTCGAAACACTGGACCCTGGGGCCTC 6900

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              |||||||:||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db       6901 CTCAGCCRATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 6960

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               ||||||||||||||||||||||||||:||||||||||||| |||||||||||||||||
Db       6961 CTACTCCTCTTTGGACCCTGTATCTTTRACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 7020

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||:|| ||||| ||||      | ||| |  || |
Db       7021 GAAGCTGTRAAACTACAAATGGAGCCCAAGATGCAGTCCAA 7061

RESULT 12
US-10-133-036-1
; Sequence 1, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
;  APPLICANT: Conrad, Bernard
;  APPLICANT:  Mach, Bernard
;  TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
;  FILE REFERENCE: 23135-507
;  CURRENT APPLICATION NUMBER: US/10/133,036
;  CURRENT FILING DATE:  2002-04-26
;  PRIOR APPLICATION NUMBER: PCT/EP00/10659
;  PRIOR FILING DATE: 2000-10-30
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2782
;   TYPE: DNA
;   ORGANISM: Human endogenous retrovirus
US-10-133-036-1

  Query Match             88.3%;  Score 1308.2;  DB 16;  Length 2782;
  Best Local Similarity   92.7%;  Pred. No. 0;
  Matches 1373;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||| || |||||| | ||||  |||| ||||||||
Db        763 ATGGCCCTCCCTTATCATATTTTTCTCTGTAGTGTTCTTTCACCCTGTTTCACTCTCACT 822

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  ||||||||||||||| ||| |||||||||||| |||
Db        823 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTCACCCAGAGTTTCTATGGAGA 882

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        883 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 942

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        943 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1002

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1003 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1062

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1063 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1122

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||| ||| ||||
Db       1123 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCGGGGTACAT 1182

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1183 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1242

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db       1243 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1302

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1303 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1362

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1422

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1423 GGACCTCTTGTTTCCAATGTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1482

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db       1483 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1542

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1543 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1602

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1662

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db       1663 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1722

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db       1723 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1782

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
```

```
               ||||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1783 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1842

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1843 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1902

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               | ||||||||||||||||||||||||||||||  | ||||||||||||||||||||||||
Db        1903 CGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTTA 1962

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| |||||||||||||||||||||||||| |||| |||||||||||||| |||||
Db        1963 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGAA 2022

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||| ||||||||||||| ||| |||||||||||||| |||||| | || ||||||||||
Db        2023 ATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2082

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||| |||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Db        2083 CTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2142

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||| |||||||||||| ||||||||||||||||||| 
Db        2143 CTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2202

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               |||||||| || ||||| ||||      | ||| |  || |
Db        2203 GAAGCTGTGAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2243

RESULT 13
US-10-632-793-26
; Sequence 26, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
;   LENGTH: 2782
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-632-793-26

  Query Match             88.2%;  Score 1306.6;  DB 16;  Length 2782;
  Best Local Similarity   92.6%;  Pred. No. 0;
  Matches 1372;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               |||| |||||||||||||| |||||||| || |||||| | ||||  |||| ||||||||
Db         763 ATGGGCCTCCCTTATCATATTTTTCTCTGTAGTGTTCTTTCACCCTGTTTCACTCTCACT 822

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  ||||||||||||||||| ||| |||||||||| |||
Db         823 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTCACCCAGAGTTTCTATGGAGA 882

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| |||||||||| | |||||||||| | ||
Db         883 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 942

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         943 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1002

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        1003 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1062

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  ||||||||||||||||||||| ||
Db        1063 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1122

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| ||||||||||||||||||| ||||| |||| ||||||||||||| ||| ||| ||||
Db        1123 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCGGGGTACAT 1182

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        1183 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1242

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| ||||||||||||||||||||||||| |||||||||||||||| |||
Db        1243 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1302

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||| ||||||||||||||  ||||||||
Db        1303 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1362

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1363 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1422

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        1423 GGACCTCTTGTTTCCAATGTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1482

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||||||||| | |||||||||| |||||||||||||||||||||||| |||||||||| |
Db        1483 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1542

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        1543 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1602

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1603 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1662

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||| | ||||| || ||||||||| ||||||||||||||||||||
Db        1663 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1722

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   |||||||||||||||||||| |||||
Db        1723 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1782

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1783 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1842

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1843 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1902

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               | ||||||||||||||||||||||||||||||  | ||||||||||||||||||||||||
Db        1903 CGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTTA 1962

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| |||||||||||||||||||||||||| |||| |||||||||||||| |||||
Db        1963 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGAA 2022
```

```
Qy      1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
             |||| ||||||||||||| ||| |||||||||||||| |||||| | || ||||||||||
Db      2023 ATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2082

Qy      1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
             ||||||| ||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db      2083 CTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2142

Qy      1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
              |||||||||||||||||||||||||| ||||||||||||| |||||||||||||||||
Db      2143 CTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2202

Qy      1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
             |||||||| || ||||| ||||      | ||| |  || |
Db      2203 GAAGCTGTGAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2243


RESULT 14
US-10-087-192-730/c
; Sequence 730, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 730
;   LENGTH: 161334
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(161334)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-730

  Query Match             86.3%;  Score 1277.4;  DB 13;  Length 161334;
  Best Local Similarity   92.5%;  Pred. No. 0;
  Matches 1399;  Conservative    0;  Mismatches   81;  Indels   33;  Gaps    4;

Qy         1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
             ||||||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Db    102037 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCCCTTTCACTCTCACT 101978

Qy        61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
             ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    101977 GCACCCCCTCCATGCTGCTGTACAACCAGCAGCTCCCCTTACCAAGAGTTTCTATGAAGA 101918

Qy       121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
             | |||||||||  |||||||||||||||||||| ||||||||||| |||| |||||||||
Db    101917 ATGCGGCTTCCCAGAAATATTGATGCCCCATCAAATAGGAGTTTACCTAAAGGAAACTCC 101858

Qy       181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
             |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    101857 ACCTTCACTGCCCACACCCATATGCCCCACAACTGCTATAACTCTGCCACTCTTTGCATG 101798

Qy       241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    101797 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAAGACTT 101738

Qy       301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
             ||||||||||||||| |||||||||||||||||||  |||||||||| ||||||||| ||
Db    101737 GGAGCCACTGTCTGTCGGACTTACTTCACCCATACTGGTATGTCTGAGGGGGGTGGAGTT 101678

Qy       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
             |||| |||||||||||||||||| |||||||||| |||||||||||||||||||| ||||
Db    101677 CAAGATCAGGCAAGAGAAAAACATGTAAAGGAAGTAATCTCCCAACTGACCCGGGTACAT 101618

Qy       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTC-AAAACTACATGAAACCCTCCGTAC 479
             |||||||||||||||||||||||||||| |||||| |||||||||||||||||||| |||
Db    101617 AGCACCCCTAGCCCCTACAAAGGACTAGATCTCTCTAAAACTACATGAAACCCTCCATAC 101558

Qy       480 CCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGC 539
             ||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||| ||
Db    101557 CCATACTTGCCTGGTAAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCGGC 101498

Qy       540 CCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAAT 599
             |||||||||||||||||||||||||||||||||||||| ||||||||||| |||||||||
Db    101497 CCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGTACTTCAGGCCATGCATTTCAAT 101438

Qy       600 CCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGT 659
             |||||| ||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    101437 CCCTGTACCTGAACAATGGAACAACTACAGCACAGAAATAAACACCACTTCCGTTTTAGT 101378

Qy       660 AGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATT 719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101377 AGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATT 101318

Qy       720 TAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACG 779
             |||||||||| |||||||||||| |||||||||||||||||||||||| |||||||||||
Db    101317 TAGCAATACTGTAGACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACG 101258

Qy       780 AATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTT 839
             |||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||
Db    101257 AATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTTAGCCTATCGTTGTTT 101198

Qy       840 GAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTA 899
             |||||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||
Db    101197 GAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCC-ATGACCATTTA 101139

Qy       900 CACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCAT 959
             |||||||||||||||||||||| |||| |||||||||||||||||||||||||||| |||
Db    101138 CACTGAACAAGATTTATACAATTATGTTGTACCTAAGCCCCACAACAAAAGAGTACTCAT 101079

Qy       960 TCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTAT 1019
             |||||||||||||||| ||||||||||||||||  |||||||| ||||||||||| |||
Db    101078 TCTTCCTTTTGTTATCGGAGCAGGAGTGCTAGGTGGACTAGGTTCTGGCATTGGCGGTAC 101019

Qy      1020 CACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACA 1079
             |||||||||||||||||||||||||||||||||||||||| | |||||||||||||||
Db    101018 CACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTCAATGGTGACATGGAATG 100959

Qy      1080 GGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCT 1139
             |||  | |||||||||||||||||||||||||||||||||| ||||||| ||||||||||
Db    100958 GGTTGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTTCCTAGCATCAGTAGTCCT 100899

Qy      1140 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 1199
             ||||||| ||||||||||||||||||||||| |  |||||||||||| ||||||||||||
Db    100898 TCAAAATTGAAGAGCTTTAGACTTGCTAACCTCTGAAAGAGGGGGAAGCTGTTTATTTTT 100839

Qy      1200 AGGAGAAGAACGCTG------------------------------TTATTATGTTAATCA 1229
             ||| |||||| | ||                              |||||||||||||||
Db    100838 AGGGGAAGAATGTTGTTATTATGTTATTTTAGCGGAAGAATGTTGTTATTATGTTAATCA 100779

Qy      1230 ATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGA 1289
             |||| |||||||||| ||||||||| |||||||||||||||| |||||||| ||||| ||||
Db    100778 ATCCTGAATTGTCACAGAGAAAGTTGAAGAAATTCGAGATTGAATACAACGTAGAACAGA 100719

Qy      1290 GGAGCTTC-AAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCC 1348
             |||||||| ||||||||  || ||||||||||||||||||||||||||||||| ||||||
Db    100718 GGAGCTTCAAAAACACCAGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCC 100659

Qy      1349 CCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTA 1408
             ||||||||||| |||||||||||||||||||| |||||||||||||||||||||||||||
Db    100658 CCTTCTTAGGATCTCTAGCAGCTCTAATATTGATACTCCTCTTTGGACCCTGTATCTTTA 100599
```

```
Qy       1409 ACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTAC 1468
              |||||||||||||||||||||||||||||||  ||| |||||||||||| || ||  | |
Db     100598 ACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGTTGTAAAGCTACAAATCGTTCTTC 100539

Qy       1469 AAATGGAACCCCA 1481
              |||||||||||||
Db     100538 AAATGGAACCCCA 100526

RESULT 15
US-09-864-761-4444
; Sequence 4444, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4444
;   LENGTH: 1894
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC002346.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =   6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444

  Query Match             76.4%;  Score 1131.6;  DB 9;  Length 1894;
  Best Local Similarity   92.4%;  Pred. No. 0;
  Matches 1218;  Conservative    0;  Mismatches   84;  Indels   16;  Gaps    2;

Qy        164 TATCTAAGGGAAACTCCACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACT 223
              |||||||||||| | |||||||||||||||||||||||||||||| ||||||||||||||
Db          1 TATCTAAGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCCACAACTGCTATAACT 60

Qy        224 CTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTA 283
              ||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||| |
Db         61 CTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAACGATTAATCCCA 120

Qy        284 GTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGT 343
              ||||||||||||||||||||             ||||| ||||||| |||||||||||||
Db        121 GTTGTCCTGGAGGACTTGGA-------------GGACTCACTTCACTCATACCAGTATGT 167

Qy        344 CTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCC 403
              |||||||||||||| |||||| |||||||| |||||||||  ||||||||| ||||||||
Db        168 CTGATGGGGGTGGAGTTCAAGATCAGGCAACAGAAAAACACATAAAGGAAGTAATCTCCC 227

Qy        404 AACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTAC 463
              |||||||| ||| ||||||||||||| |||||||||||||||||| ||||||||||||||
Db        228 AACTGACCTGGGTACATAGCACCCCTGGCCCCTACAAAGGACTAGATCTCTCAAAACTAC 287

Qy        464 ATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGC 523
              |||||||||||| |||||||||| ||||||| |||||||||||||||||||| ||| |||
Db        288 ATGAAACCCTCCATACCCATACTGGCCTGGTAAGCCTATTTAATACCACCCTGACTGGGC 347

Qy        524 TCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCA 583
              ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||| |
Db        348 TCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTTA 407

Qy        584 GGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACA 643
              |||||||||||||||||||| | |||||||||||||||||||||||||||||||||||||
Db        408 GGCCATACATTTCAATCCCTATACCTGAACAATGGAACAACTTCAGCACAGAAATAAACA 467

Qy        644 CCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACC 703
              ||||||| ||||||||||| ||||   |||||||||||||||||||||||||||||||||
Db        468 CCACTTCTGTTTTAGTAGGTCCTC---TTTCCAATCTGGAAATAACCCATACCTCAAACC 524

Qy        704 TCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGG 763
              ||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db        525 TCACCTGTGTAAAATTTAGCAATACTATAGACACAGCCAACTCCCAATGCATCAGGTGGG 584

Qy        764 TAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCT 823
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 TAACTCCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCT 644

Qy        824 CAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGC 883
              ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        645 CAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTGTGTGCTTCCTCTCATTCTTAGTGG 704

Qy        884 CCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACA 943
              |||||||| ||||||||||||||||||||||||||||||||||| |||||||||||| ||
Db        705 CCCCTATGCCCATCTACACTGAACAAGATTTATACAATCATGTCATACCTAAGCCCCGCA 764

Qy        944 ACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTA 1003
              |||||||||||||||||||||||||||||||  ||||||||||||||||| || ||| ||
Db        765 ACAAAAGAGTACCCATTCTTCCTTTTGTTATTGGAGCAGGAGTGCTAGGCGGAGTAGCTA 824

Qy       1004 CTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAA 1063
              ||||||||||| |||||||||||||||||||||||||||||||||| ||||||||| |||
Db        825 CTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTGTCTCAAGAACTAA 884

Qy       1064 ATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCC 1123
              ||||||||||||||  |||| ||||  |||||||||||||||||||||||||||||||||
Db        885 ATGGTGACATGGAATGGGTCGCTGATACCCTGGTCACCTTGCAAGATCAACTTAACTCCC 944
```

```
Qy      1124 TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGG 1183
             |||||||||||||||||||||||||||||||||||||||||||||||||  |||| ||||
Db       945 TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGG 1004

Qy      1184 GAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCA 1243
             |||||| |||||||||||  ||| || |||||| |||||||||||||||| ||||  |||
Db      1005 GAACCTTTTTATTTTTAGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCA 1064

Qy      1244 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACA 1303
             | ||||||||||||||||||| || ||||||| || |||||||| ||||||| |||||||
Db      1065 CCGAGAAAGTTAAAGAAATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACA 1124

Qy      1304 CCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTC 1363
             | | || ||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      1125 CTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTC 1184

Qy      1364 TAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGT 1423
             |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1185 TAGCAGCTATAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGT 1244

Qy      1424 TTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
             |||||| ||||||||| ||||| ||||| ||||| || ||  | |||||||| |||||
Db      1245 TTGTCTTTTCCAGAATCGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCA 1302


RESULT 16
US-10-632-793-24
; Sequence 24, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 1948
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84)..(84)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (193)..(193)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (241)..(241)
;   OTHER INFORMATION: n = a or g or c or t/u
US-10-632-793-24

  Query Match             75.4%;  Score 1116.4;  DB 16;  Length 1948;
  Best Local Similarity   94.2%;  Pred. No. 0;
  Matches 1159;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
             ||||||||||||||||||| |||||||||||||||| | | |||| ||||| ||||||||
Db       719 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTGTTTCACCCTCTTTCACTCTCACT 778

Qy        61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
             ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db       779 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 838

Qy       121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
             | || || ||| |||||||||||||||||||| ||||||||| | | |||||||| | ||
Db       839 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTGTAAGGGAACCCCC 898

Qy       181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
             ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       899 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 958

Qy       241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
             |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       959 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1018

Qy       301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
             |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db      1019 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1078

Qy       361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
             |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db      1079 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1138

Qy       421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
              ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      1139 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1198

Qy       481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
             |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db      1199 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1258

Qy       541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
             ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db      1259 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1318

Qy       601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
             ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1319 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1378

Qy       661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1379 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1438

Qy       721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
             |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db      1439 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1498

Qy       781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      1499 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1558

Qy       841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db      1559 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGGCCATCTAC 1618

Qy       901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
             ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db      1619 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1678

Qy       961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
             ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db      1679 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1738

Qy      1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
             ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db      1739 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1798

Qy      1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
             ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1799 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1858

Qy      1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGAACCTGTTTATTTTTA 1200
```

```
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1859 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1918

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAA 1230
              || |||||| ||||||||||||||||||||
Db       1919 GGGGAAGAATGCTGTTATTATGTTAATCAA 1948


RESULT 17
US-09-864-761-21192
; Sequence 21192, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21192
;   LENGTH: 792
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC002346.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =   6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
;   OTHER INFORMATION: NT HIT: AF208161.1, EVALUE 0.00e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AU138405.1, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: P10269, EVALUE 2.00e-16
US-09-864-761-21192

  Query Match             48.4%;  Score 716.6;  DB 9;  Length 792;
  Best Local Similarity   94.7%;  Pred. No. 1e-209;
  Matches  753;  Conservative    0;  Mismatches   39;  Indels    3;  Gaps    1;

Qy        375 AGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCC 434
              |||||||||  ||||||||| |||||||||||||||| ||| ||||||||||||| ||||
Db          1 AGAAAAACACATAAAGGAAGTAATCTCCCAACTGACCTGGGTACATAGCACCCCTGGCCC 60

Qy        435 CTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGT 494
              |||||||||||||| |||||||||||||||||||||||||| |||||||||| |||||||
Db         61 CTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCATACCCATACTGGCCTGGT 120

Qy        495 GAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAA 554
               |||||||||||||||||||| ||| |||||||||||||||| |||||||||||||||||
Db        121 AAGCCTATTTAATACCACCCTGACTGGGCTCCATGAGGTCTCGGCCCAAAACCCTACTAA 180

Qy        555 CTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACA 614
              ||||||||||||||||||||||||||| ||||||||||||||||||||| | ||||||||
Db        181 CTGTTGGATGTGCCTCCCCCTGCACTTTAGGCCATACATTTCAATCCCTATACCTGAACA 240

Qy        615 ATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTC 674
              |||||||||||||||||||||||||||||||||||| ||||||||||| ||||   ||||
Db        241 ATGGAACAACTTCAGCACAGAAATAAACACCACTTCTGTTTTAGTAGGTCCTC---TTTC 297

Qy        675 CAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGA 734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 CAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGA 357

Qy        735 CACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACC 794
              |||| ||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        358 CACAGCCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACGAATAGTCTGCCTACC 417

Qy        795 CTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGA 854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        418 CTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGA 477

Qy        855 ATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTT 914
              |||| ||||||||||||||||||||||| |||||||| ||||||||||||||||||||||
Db        478 ATCTGTGTGCTTCCTCTCATTCTTAGTGGCCCCTATGCCCATCTACACTGAACAAGATTT 537

Qy        915 ATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTAT 974
              ||||||||||||| |||||||||||| |||||||||||||||||||||||||||||||||
Db        538 ATACAATCATGTCATACCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTTAT 597

Qy        975 CAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCA 1034
                ||||||||||||||||| || ||| ||||||||||||| |||||||||||||||||||
Db        598 TGGAGCAGGAGTGCTAGGCGGAGTAGCTACTGGCATTGGCGGTATCACAACCTCTACTCA 657

Qy       1035 GTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCT 1094
              ||||||||||||||| ||||||||| |||||||||||||||||  |||| ||||  ||||
Db        658 GTTCTACTACAAACTGTCTCAAGAACTAAATGGTGACATGGAATGGGTCGCTGATACCCT 717

Qy       1095 GGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGC 1154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGC 777

Qy       1155 TTTAGACTTGCTAAC 1169
              |||||||||||||||
Db        778 TTTAGACTTGCTAAC 792


RESULT 18
US-10-363-616-228
; Sequence 228, Application US/10363616
; Publication No. US20040044181A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 21272-113 (793)
;  CURRENT APPLICATION NUMBER: US/10/363,616
;  CURRENT FILING DATE:  2003-03-03
;  PRIOR APPLICATION NUMBER: 09/654,935
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 490
; SEQ ID NO 228
;   LENGTH: 1684
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (89)..(1684)
US-10-363-616-228

  Query Match             40.0%;  Score 591.8;  DB 16;  Length 1684;
  Best Local Similarity   98.8%;  Pred. No. 4.8e-171;
  Matches  596;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        879 AGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCC 938
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCC 780

Qy        939 CCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACT 998
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACT 840

Qy        999 AGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGA 1058
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGA 900

Qy       1059 AATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAA 1118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAA 960

Qy       1119 CTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAG 1178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAG 1020

Qy       1179 AGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAAT 1238
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1021 AGGGGGAACCTGTTTATTTTTAGGAGAAGAATGCTGTTATTATGTTAATCAATCCAGAAT 1080

Qy       1239 TGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCA 1298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCA 1140

Qy       1299 AAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGG 1358
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAACACCGAACACTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGG 1200

Qy       1359 ACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGT 1418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGT 1260

Qy       1419 TAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACC 1478
              ||||||||||||||||||||||||||||||||||||||| |||||  | |||||||| ||
Db       1261 TAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAAATGGTTCTTCAAATGGAGCC 1320

Qy       1479 CCA 1481
              |||
Db       1321 CCA 1323

RESULT 19
US-10-632-793-19
; Sequence 19, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
```

```
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-632-793-19

  Query Match             39.4%;  Score 582.8;  DB 16;  Length 591;
  Best Local Similarity   99.7%;  Pred. No. 1.6e-168;
  Matches  584;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        887 CTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACA 946
              | ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCATGGCCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACA 60

Qy        947 AAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTG 1006
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTG 120

Qy       1007 GCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATG 1066
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATG 180

Qy       1067 GTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAG 1126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAG 240

Qy       1127 CAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAA 1186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAA 300

Qy       1187 CCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTG 1246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTG 360

Qy       1247 AGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCG 1306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCG 420

Qy       1307 AACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAG 1366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAG 480

Qy       1367 CAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTG 1426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTG 540

Qy       1427 TCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAAT 1472
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAAT 586

RESULT 20
US-10-632-793-25
; Sequence 25, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 1136
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-632-793-25

  Query Match             34.9%;  Score 516.6;  DB 16;  Length 1136;
  Best Local Similarity   91.8%;  Pred. No. 5.9e-148;
  Matches  546;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         887 CTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACA 946
               | ||| ||||||||||||||||||||||||||| | ||||| || ||||||||| |||||
Db           1 CCATGGCCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACA 60

Qy         947 AAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTG 1006
               ||||||||||||||||||||||||||||  ||||||||||||||||   |||||||||||
Db          61 AAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTG 120

Qy        1007 GCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATG 1066
               |||||||| |||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db         121 GCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATG 180

Qy        1067 GTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAG 1126
               | |||||||||| |||| | ||||||||||||||||||||||||||||||||||||||||
Db         181 GGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAG 240

Qy        1127 CAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAA 1186
               |||||||||||||||||||||||||||||||||||| |||||||||  ||||||||||||
Db         241 CAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTCGCTAACCGCTGAAAGAGGGGGAA 300

Qy        1187 CCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTG 1246
               |||||||||||||||| |||||| ||||||||||||||||||||||| |||| |||||||
Db         301 CCTGTTTATTTTTAGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTG 360

Qy        1247 AGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCG 1306
               |||||||||||||||||||||||||||||||| |||||||||| ||||||| |||||| |
Db         361 AGAAAGTTAAAGAAATTCGAGATCGAATACAACGTAGAGCAGAAGAGCTTCGAAACACTG 420

Qy        1307 AACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAG 1366
                || ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db         421 GACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAG 480

Qy        1367 CAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTG 1426
               ||||| |||||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db         481 CAGCTATAATATTGCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTG 540

Qy        1427 TCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||||| ||||||||||| ||||| ||||      | ||| |  || |
Db         541 TCTCTTCCAGAATCGAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 595

RESULT 21
US-10-632-793-23
; Sequence 23, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
```

```
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
;   LENGTH: 2006
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (305)..(305)
;   OTHER INFORMATION: n = a or g or c or t/u
US-10-632-793-23

  Query Match             34.7%;  Score 513.6;  DB 16;  Length 2006;
  Best Local Similarity   93.2%;  Pred. No. 6.9e-147;
  Matches  548;  Conservative    0;  Mismatches   39;  Indels    1;  Gaps    1;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        1420 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 1479

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        1480 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 1539

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        1540 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 1599

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        1600 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1659

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        1660 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1719

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        1720 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1779

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        1780 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1839

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| |||| |||||||||||||||| |||||||||||||||||||||||||||||||
Db        1840 GGCACCTCTAG-CCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1898

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        1899 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1958

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCA 588
               ||||||||||||||||||||||| |||||||||||| |||||| ||||
Db        1959 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAAGCCA 2006

RESULT 22
US-10-027-632-322491/c
; Sequence 322491, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 322491
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(570)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322491

  Query Match             25.5%;  Score 377.8;  DB 13;  Length 570;
  Best Local Similarity   90.7%;  Pred. No. 2.6e-105;
  Matches  411;  Conservative    2;  Mismatches   39;  Indels    1;  Gaps    1;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              || |||||||||||||||| ||||||||||||  ||||||||||  ||||| ||||||||
Db        463 ATAGCCCTCCCTTATCATATTTTTCTCTTTACCATTCTCTTACCTGCTTTCACTCTCACT 404

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||| | ||||||||||:||||  |||||||||||||||||||| |||||||| || |||
Db        403 TCACTCACTCCATGCTGYTGTATGACCAGTAGCTCCCCTTACCACGAGTTTCT-TGGAGA 345

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || ||||||||||||||||||| |||||||||||||||||||||||||||||||   |
Db        344 ATGCAGCTTCCTGGAAATATTGATTCCCCATCATATAGGAGTTTATCTAAGGGAAATCAC 285

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              |||||||||||||||||||||||||||  ||||||||||||| ||:||||||||||||||
Db        284 ACCTTCACTGCCCACACCCATATGCCCTACAACTGCTATAACGCTRCCACTCTTTGCATG 225

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||| |||||||  || ||||||||||||||||| |||||||||||||||||
Db        224 CATGCAAATACTAATTATTGTTCATGGAAAATGATTAATCCTTGTTGTCCTGGAGGACTT 165

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||| ||
Db        164 GGAGCCACTGTCTGTTGGACTTACTTCACCCCTACCAGTATGTCTGATGGGGGTGGAGTT 105

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| |||||||||| ||||||||||||||||| || ||||
Db        104 CAAGATCAGGCAAGAGAAAAACACGTAAAGGAAGTAATCTCCCAACTGACCCTGGTACAT 45

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTC 453
              ||||||||||| ||||||||||||||| | | |
Db         44 AGCACCCCTAGACCCTACAAAGGACTAATCCCC 12

RESULT 23
US-10-027-632-322492/c
; Sequence 322492, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 322492
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(570)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322492

  Query Match             25.5%;  Score 377.8;  DB 13;  Length 570;
  Best Local Similarity   90.7%;  Pred. No. 2.6e-105;
  Matches  411;  Conservative    2;  Mismatches   39;  Indels    1;  Gaps    1;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              || |||||||||||||||| ||||||||||||  ||||||||||  ||||| ||||||||
Db        463 ATAGCCCTCCCTTATCATATTTTTCTCTTTACCATTCTCTTACCTGCTTTCACTCTCACT 404

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||| | ||||||||||:||||  |||||||||||||||||||| |||||||| || |||
Db        403 TCACTCACTCCATGCTGYTGTATGACCAGTAGCTCCCCTTACCACGAGTTTCT-TGGAGA 345

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || ||||||||||||||||||| |||||||||||||||||||||||||||||||   |
Db        344 ATGCAGCTTCCTGGAAATATTGATTCCCCATCATATAGGAGTTTATCTAAGGGAAATCAC 285

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              |||||||||||||||||||||||||||  ||||||||||||| ||:||||||||||||||
Db        284 ACCTTCACTGCCCACACCCATATGCCCTACAACTGCTATAACGCTRCCACTCTTTGCATG 225

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||| |||||||  || ||||||||||||||||| |||||||||||||||||
Db        224 CATGCAAATACTAATTATTGTTCATGGAAAATGATTAATCCTTGTTGTCCTGGAGGACTT 165

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||| ||
Db        164 GGAGCCACTGTCTGTTGGACTTACTTCACCCCTACCAGTATGTCTGATGGGGGTGGAGTT 105

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| |||||||||| ||||||||||||||||| || ||||
Db        104 CAAGATCAGGCAAGAGAAAAACACGTAAAGGAAGTAATCTCCCAACTGACCCTGGTACAT 45

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTC 453
              ||||||||||| ||||||||||||||| | | |
Db         44 AGCACCCCTAGACCCTACAAAGGACTAATCCCC 12

RESULT 24
US-10-027-632-322491/c
; Sequence 322491, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 322491
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(570)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322491

  Query Match             25.5%;  Score 377.8;  DB 15;  Length 570;
  Best Local Similarity   90.7%;  Pred. No. 2.6e-105;
  Matches  411;  Conservative    2;  Mismatches   39;  Indels    1;  Gaps    1;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               || |||||||||||||||| ||||||||||||  ||||||||||  ||||| ||||||||
Db         463 ATAGCCCTCCCTTATCATATTTTTCTCTTTACCATTCTCTTACCTGCTTTCACTCTCACT 404

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
                ||| | ||||||||||:||||  ||||||||||||||||||||| |||||||| || |||
Db         403 TCACTCACTCCATGCTGYTGTATGACCAGTAGCTCCCCTTACCACGAGTTTCT-TGGAGA 345

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || |||||||||||||||||||| |||||||||||||||||||||||||||||   |
Db         344 ATGCAGCTTCCTGGAAATATTGATTCCCCATCATATAGGAGTTTATCTAAGGGAAATCAC 285

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               |||||||||||||||||||||||||||  ||||||||||||| ||:||||||||||||||
Db         284 ACCTTCACTGCCCACACCCATATGCCCTACAACTGCTATAACGCTRCCACTCTTTGCATG 225

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||| |||||||  || ||||||||||||||||||| ||||||||||||||||
Db         224 CATGCAAATACTAATTATTGTTCATGGAAAATGATTAATCCTTGTTGTCCTGGAGGACTT 165

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db         164 GGAGCCACTGTCTGTTGGACTTACTTCACCCCTACCAGTATGTCTGATGGGGGTGGAGTT 105

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| |||||||||| ||||||||||||||||||| || ||||
Db         104 CAAGATCAGGCAAGAGAAAAACACGTAAAGGAAGTAATCTCCCAACTGACCCTGGTACAT 45

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTC 453
               ||||||||||| ||||||||||||||||| | | |
Db          44 AGCACCCCTAGACCCTACAAAGGACTAATCCCC 12

RESULT 25
US-10-027-632-322492/c
```

```
; Sequence 322492, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 322492
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(570)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322492

  Query Match             25.5%;  Score 377.8;  DB 15;  Length 570;
  Best Local Similarity   90.7%;  Pred. No. 2.6e-105;
  Matches  411;  Conservative    2;  Mismatches   39;  Indels    1;  Gaps    1;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               || |||||||||||||||| ||||||||||||  ||||||||||  ||||| ||||||||
Db         463 ATAGCCCTCCCTTATCATATTTTTCTCTTTACCATTCTCTTACCTGCTTTCACTCTCACT 404

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
                ||| | ||||||||||:||||  ||||||||||||||||||||| |||||||| || |||
Db         403 TCACTCACTCCATGCTGYTGTATGACCAGTAGCTCCCCTTACCACGAGTTTCT-TGGAGA 345

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || |||||||||||||||||||| |||||||||||||||||||||||||||||   |
Db         344 ATGCAGCTTCCTGGAAATATTGATTCCCCATCATATAGGAGTTTATCTAAGGGAAATCAC 285

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               |||||||||||||||||||||||||||  ||||||||||||| ||:||||||||||||||
Db         284 ACCTTCACTGCCCACACCCATATGCCCTACAACTGCTATAACGCTRCCACTCTTTGCATG 225

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||| |||||||  || ||||||||||||||||||| ||||||||||||||||
Db         224 CATGCAAATACTAATTATTGTTCATGGAAAATGATTAATCCTTGTTGTCCTGGAGGACTT 165

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db         164 GGAGCCACTGTCTGTTGGACTTACTTCACCCCTACCAGTATGTCTGATGGGGGTGGAGTT 105

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| |||||||||| ||||||||||||||||||| || ||||
Db         104 CAAGATCAGGCAAGAGAAAAACACGTAAAGGAAGTAATCTCCCAACTGACCCTGGTACAT 45

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTC 453
               ||||||||||| ||||||||||||||||| | | |
Db          44 AGCACCCCTAGACCCTACAAAGGACTAATCCCC 12

RESULT 26
```

```
US-10-029-386-4312/c
; Sequence 4312, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4312
;   LENGTH: 521
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC021619.3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =   6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.81
;   OTHER INFORMATION: EST_HUMAN HIT: BE734284.1, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: P10269, EVALUE 4.00e-26
;   OTHER INFORMATION: NT HIT: AF208161.1, EVALUE 0.00e+00
US-10-029-386-4312

  Query Match            24.6%;  Score 364;  DB 15;  Length 521;
  Best Local Similarity  91.7%;  Pred. No. 4.4e-101;
  Matches  385;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     1062 AAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTC 1121
            ||||||||||||||||  |||||| |||||||||||||||||||||||||||||||||||
Db      521 AAATGGTGACATGGAATGGGTCACCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTC 462

Qy     1122 CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGG 1181
            | |||||||||||||||||||||||||||||||||||||||||||||| ||| |||||||
Db      461 CTTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACTGCCGAAAGAGG 402

Qy     1182 GGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGT 1241
            ||||| ||||||||||||||  |||||| |||||||||| |||||||||||  |||| ||
Db      401 GGGAATCTGTTTATTTTTAGTGGAAGAATGCTGTTATTACGTTAATCAATCTGGAATCGT 342

Qy     1242 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAA 1301
            ||| || |||||||||||||||| |||||||||||||  |||||||||||||||||||||
Db      341 CACCGAAAAAGTTAAAGAAATTCAAGATCGAATACAACATAGAGCAGAGGAGCTTCAAAA 282

Qy     1302 CACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACC 1361
            ||||  || ||||| ||||||||||||||||||||||||| |||||||||||||||||||
Db      281 CACCAGACCCTGGGACCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACC 222

Qy     1362 TCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAA 1421
            ||||||  || ||||| |||||||||||||||||||||||||||||||||||||| ||||
Db      221 TCTAGCGACTATAATACTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTTTTAA 162

Qy     1422 GTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
            |||||||||||||||||||||||||||||||||||| ||||   | |||||||| |||||
Db      161 GTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAAATGGCTCTTCAAATGGAGCCCCA 102


RESULT 27
US-10-027-632-322574
; Sequence 322574, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 322574
;   LENGTH: 551
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(551)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322574

  Query Match            23.8%;  Score 353;  DB 13;  Length 551;
  Best Local Similarity  92.0%;  Pred. No. 1.1e-97;
  Matches  412;  Conservative    2;  Mismatches   30;  Indels    4;  Gaps    4;

Qy        1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
            ||||||||||||||||||| ||||||||||||||||||||||||| ||||| ||||||||
Db      102 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCTCTTTCACTCTCACT 161

Qy       61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
            |||||| |||||||| ||||||| ||||||||||||||||||||| ||||||||||:|||
Db      162 GCACCCGCTCCATGCCGCTGTACGACCAGTAGCTCCCCTTACCAATAGTTTCTATGRAGA 221

Qy      121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
            | |||||||||||||||||||||||||||||||:||||||||||||||||||||||| ||
Db      222 ATGCGGCTTCCTGGAAATATTGATGCCCCATCAWATAGGAGTTTATCTAAGGGAAAC-CC 280

Qy      181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
            |||||||||||||||||||||||||||| ||| |||| ||||||||||||||||||||||
Db      281 ACCTTCACTGCCCACACCCATATGCCCCACAAATGCTGTAACTCTGCCACTCTTTGCATG 340

Qy      241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db      341 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTTGAGGACTT 400

Qy      301 GGAGCCACTGTCTGTTGGACTT-ACTTCACCCATACCAGTATGTCTGATGGGGGTGGAAT 359
            |||||||||||||||||||||| ||||||| |||||| |||||||||||||||||||| |
Db      401 GGAGCCACTGTCTGTTGGACTTAACTTCACTCATACCGGTATGTCTGATGGGGGTGGAGT 460

Qy      360 TCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACA 419
            || || | ||||||||| |||||| ||| |||||| ||    ||| |||||   || |||
Db      461 TC-AGATGAGGCAAGAG-AAAACACGTANAGGAAGTAAATCTCCANCTGACNGTGGTACA 518

Qy      420 TAGCACCCCTAGCCCCTACAAAGGACTA 447
            ||||||||||| ||||||||||||||||
Db      519 TAGCACCCCTAACCCCTACAAAGGACTA 546


RESULT 28
US-10-027-632-322575
; Sequence 322575, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 322575
;   LENGTH: 551
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(551)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322575

  Query Match             23.8%;  Score 353;  DB 13;  Length 551;
  Best Local Similarity   92.0%;  Pred. No. 1.1e-97;
  Matches  412;  Conservative    2;  Mismatches   30;  Indels    4;  Gaps    4;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| ||||||||||||||||||||||||| ||||| ||||||||
Db        102 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCTCTTTCACTCTCACT 161

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              |||||| |||||||| ||||||| ||||||||||||||||||||| ||||||||||:|||
Db        162 GCACCCGCTCCATGCCGCTGTACGACCAGTAGCTCCCCTTACCAATAGTTTCTATGRAGA 221

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | |||||||||||||||||||||||||||||||:||||||||||||||||||||||| ||
Db        222 ATGCGGCTTCCTGGAAATATTGATGCCCCATCAWATAGGAGTTTATCTAAGGGAAAC-CC 280

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              |||||||||||||||||||||||||||| ||| |||| ||||||||||||||||||||||
Db        281 ACCTTCACTGCCCACACCCATATGCCCCACAAATGCTGTAACTCTGCCACTCTTTGCATG 340

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        341 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTTGAGGACTT 400

Qy        301 GGAGCCACTGTCTGTTGGACTT-ACTTCACCCATACCAGTATGTCTGATGGGGGTGGAAT 359
              |||||||||||||||||||||| ||||||| |||||| |||||||||||||||||||| |
Db        401 GGAGCCACTGTCTGTTGGACTTAACTTCACTCATACCGGTATGTCTGATGGGGGTGGAGT 460

Qy        360 TCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACA 419
              || || | ||||||||| |||||| ||| |||||| ||    ||| |||||   || |||
Db        461 TC-AGATGAGGCAAGAG-AAAACACGTANAGGAAGTAAATCTCCANCTGACNGTGGTACA 518

Qy        420 TAGCACCCCTAGCCCCTACAAAGGACTA 447
              ||||||||||| ||||||||||||||||
Db        519 TAGCACCCCTAACCCCTACAAAGGACTA 546

RESULT 29
US-10-027-632-322574
; Sequence 322574, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
```

```
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 322574
;   LENGTH: 551
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(551)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322574

  Query Match             23.8%;  Score 353;  DB 15;  Length 551;
  Best Local Similarity   92.0%;  Pred. No. 1.1e-97;
  Matches  412;  Conservative    2;  Mismatches   30;  Indels    4;  Gaps    4;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| ||||||||||||||||||||||||| ||||| ||||||||
Db        102 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCTCTTTCACTCTCACT 161

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              |||||| |||||||| ||||||| ||||||||||||||||||||| ||||||||||:|||
Db        162 GCACCCGCTCCATGCCGCTGTACGACCAGTAGCTCCCCTTACCAATAGTTTCTATGRAGA 221

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | |||||||||||||||||||||||||||||||:||||||||||||||||||||||| ||
Db        222 ATGCGGCTTCCTGGAAATATTGATGCCCCATCAWATAGGAGTTTATCTAAGGGAAAC-CC 280

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              |||||||||||||||||||||||||||| ||| |||| ||||||||||||||||||||||
Db        281 ACCTTCACTGCCCACACCCATATGCCCCACAAATGCTGTAACTCTGCCACTCTTTGCATG 340

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        341 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTTGAGGACTT 400

Qy        301 GGAGCCACTGTCTGTTGGACTT-ACTTCACCCATACCAGTATGTCTGATGGGGGTGGAAT 359
              |||||||||||||||||||||| ||||||| |||||| |||||||||||||||||||| |
Db        401 GGAGCCACTGTCTGTTGGACTTAACTTCACTCATACCGGTATGTCTGATGGGGGTGGAGT 460

Qy        360 TCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACA 419
              || || | ||||||||| |||||| ||| |||||| ||    ||| |||||   || |||
Db        461 TC-AGATGAGGCAAGAG-AAAACACGTANAGGAAGTAAATCTCCANCTGACNGTGGTACA 518

Qy        420 TAGCACCCCTAGCCCCTACAAAGGACTA 447
              ||||||||||| ||||||||||||||||
Db        519 TAGCACCCCTAACCCCTACAAAGGACTA 546

RESULT 30
US-10-027-632-322575
; Sequence 322575, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 322575
;   LENGTH: 551
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(551)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322575

  Query Match             23.8%;  Score 353;  DB 15;  Length 551;
  Best Local Similarity   92.0%;  Pred. No. 1.1e-97;
  Matches  412;  Conservative    2;  Mismatches   30;  Indels    4;  Gaps    4;

Qy            1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
                ||||||||||||||||||| ||||||||||||||||||||||||| ||||| ||||||||
Db          102 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTCTTACCCTCTTTCACTCTCACT 161

Qy           61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
                |||||| |||||||| ||||||| ||||||||||||||||||||| ||||||||||:|||
Db          162 GCACCCGCTCCATGCCGCTGTACGACCAGTAGCTCCCCTTACCAATAGTTTCTATGRAGA 221

Qy          121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
                | |||||||||||||||||||||||||||||||:||||||||||||||||||||||| ||
Db          222 ATGCGGCTTCCTGGAAATATTGATGCCCCATCAWATAGGAGTTTATCTAAGGGAAAC-CC 280

Qy          181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
                |||||||||||||||||||||||||||| ||| |||| ||||||||||||||||||||||
Db          281 ACCTTCACTGCCCACACCCATATGCCCCACAAATGCTGTAACTCTGCCACTCTTTGCATG 340

Qy          241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          341 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTTGAGGACTT 400

Qy          301 GGAGCCACTGTCTGTTGGACTT-ACTTCACCCATACCAGTATGTCTGATGGGGGTGGAAT 359
                |||||||||||||||||||||| ||||||| |||||| |||||||||||||||||||| |
Db          401 GGAGCCACTGTCTGTTGGACTTAACTTCACTCATACCGGTATGTCTGATGGGGGTGGAGT 460

Qy          360 TCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACA 419
                || || | ||||||||| |||||| ||| |||||| ||    ||| |||||   || |||
Db          461 TC-AGATGAGGCAAGAG-AAAACACGTANAGGAAGTAAATCTCCANCTGACNGTGGTACA 518

Qy          420 TAGCACCCCTAGCCCCTACAAAGGACTA 447
                ||||||||||| ||||||||||||||||
Db          519 TAGCACCCCTAACCCCTACAAAGGACTA 546

Search completed: January 22, 2005, 20:08:34
Job time : 812.34 secs
```

```
                       GenCore version 5.1.6
             Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          January 22, 2005, 12:01:19 ; Search time 4411.62 Seconds
                                         (without alignments)
                                         12232.983 Million cell updates/sec

Title:           US-09-319-156B-9
Perfect score:   1481
Sequence:        1 atggccctcccttatcatac..........gtcttacaaatggaacccca 1481

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters:      65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                 1:  gb_est1:*
                 2:  gb_est2:*
                 3:  gb_htc:*
                 4:  gb_est3:*
                 5:  gb_est4:*
                 6:  gb_est5:*
                 7:  gb_est6:*
                 8:  gb_gss1:*
                 9:  gb_gss2:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                    %
Result              Query
  No.    Score     Match Length DB  ID                      Description
 -----------------------------------------------------------------------------
     1   1333.8    90.1    2500  3  CR622175                CR622175 full-leng
     2   1333.8    90.1    2716  3  CR613169                CR613169 full-leng
     3   1333.8    90.1    2748  3  CR605851                CR605851 full-leng
     4   1333.8    90.1    2748  3  CR625046                CR625046 full-leng
     5   1333.8    90.1    2749  3  CR617248                CR617248 full-leng
c    6      839    56.7     935  5  BX391741                BX391741 BX391741
     7    813.6    54.9     955  5  BX347952                BX347952 BX347952
     8    752.2    50.8     931  5  BX389656                BX389656 BX389656
c    9    747.2    50.5     872  5  BX326647                BX326647 BX326647
    10      741    50.0     903  5  BX408752                BX408752 BX408752
    11    737.2    49.8     883  5  BX408733                BX408733 BX408733
c   12    728.4    49.2     877  5  BX347619                BX347619 BX347619
c   13      714    48.2     836  5  BX388766                BX388766 BX388766
    14    709.8    47.9     921  5  BX408734                BX408734 BX408734
c   15    705.8    47.7     844  5  BX347111                BX347111 BX347111
    16    703.2    47.5     828  5  BX367907                BX367907 BX367907
c   17    698.6    47.2     846  5  BX368078                BX368078 BX368078
    18    696.2    47.0     995  5  BX389657                BX389657 BX389657
    19      695    46.9     924  5  BX347314                BX347314 BX347314
    20      647    43.7     723  1  AU138405                AU138405 AU138405
    21    629.4    42.5     631  8  AQ261133                AQ261133 CITBI-E1-
    22    628.6    42.4     702  1  AU138097                AU138097 AU138097
    23    623.8    42.1     924  5  BX409328                BX409328 BX409328
    24    609.8    41.2     714  2  BE734284                BE734284 601565487
c   25    580.4    39.2     658  5  BX430050                BX430050 BX430050
    26    555.8    37.5     792  5  BX409304                BX409304 BX409304
c   27    537.6    36.3     814  5  BX430055                BX430055 BX430055
    28    536.4    36.2     586  2  BE019603                BE019603 ba84f03.y
    29    535.2    36.1     790  4  BG572445                BG572445 602593490
c   30    478.6    32.3     760  5  BX450495                BX450495 BX450495
c   31      444    30.0    1071  5  BX365066                BX365066 BX365066
    32    421.6    28.5     944  5  BX368759                BX368759 BX368759
    33    416.4    28.1     890  5  BX326300                BX326300 BX326300
c   34    412.2    27.8     998  5  BX337769                BX337769 BX337769
c   35    412.2    27.8    1058  5  BX378303                BX378303 BX378303
    36    411.4    27.8     905  5  BX429316                BX429316 BX429316
c   37    406.4    27.4     564  5  BX388917                BX388917 BX388917
c   38      402    27.1    1019  5  BX439636                BX439636 BX439636
c   39    401.4    27.1     484  5  BX409035                BX409035 BX409035
c   40    399.6    27.0     494  5  BQ365207                BQ365207 RC6-GN007
c   41    390.2    26.3     427  1  AA776439                AA776439 zj50h06.s
c   42    389.8    26.3     718  1  AL601847                AL601847 DKFZp313K
c   43    373.2    25.2     966  5  BX380176                BX380176 BX380176
c   44    371.4    25.1     586  8  AQ776289                AQ776289 HS_5528_A
    45    356.4    24.1     459  7  R68740                  R68740 yi14g06.r1


                                ALIGNMENTS


RESULT 1
CR622175
LOCUS       CR622175                2500 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI051YM13 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR622175
VERSION     CR622175.1  GI:50502982
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2500)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2500)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2500
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             90.1%;  Score 1333.8;  DB 3;  Length 2500;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db        530 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 589
```

```
Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         590 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 649

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db         650 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 709

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         710 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 769

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db         770 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 829

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db         830 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 889

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db         890 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 949

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         950 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1009

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        1010 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1069

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        1070 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1129

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1130 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1189

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1190 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1249

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db        1250 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1309

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        1310 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1369

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1370 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1429

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db        1430 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1489

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        1490 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1549

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1550 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1609

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1610 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1669

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        1670 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1729

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db        1730 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1789

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db        1790 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 1849

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        1850 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 1909

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db        1910 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 1969

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||| ||||| ||||      | ||| |  || |
Db        1970 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2010


RESULT 2
CR613169
LOCUS       CR613169                2716 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DE013YI20 of Placenta of Homo sapiens
            (human).
ACCESSION   CR613169
VERSION     CR613169.1  GI:50493976
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2716)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2716)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2716
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE013YI20"
                     /tissue_type="Placenta"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             90.1%;  Score 1333.8;  DB 3;  Length 2716;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db         773 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 832
```

```
Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||||| |||
Db        833 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 892

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        893 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 952

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        953 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1012

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1013 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1072

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db       1073 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1132

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       1133 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1192

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1193 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1252

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| ||||||||||||||||||||||||| |||||||||||||||| |||
Db       1253 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1312

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1313 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1372

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1373 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTAGTA 1432

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1433 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1492

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              |||||||||| | |||||||||| |||||||||||||||||||||||||| |||||||| |
Db       1493 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1552

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1553 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1612

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1613 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1672

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db       1673 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1732

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||||   |||||||||||||||||| |||||
Db       1733 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1792

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1793 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1852

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1853 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1912

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1913 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1972

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || |||||| |||||||||||||||||||||||| |||| ||||||||||||||||||||
Db       1973 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2032

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              ||||||||||||||||| ||||||||||||||||||| |||||| | || ||||||||||
Db       2033 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2092

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              ||||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db       2093 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2152

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       2153 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2212

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  ||  |
Db       2213 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2253


RESULT 3
CR605851
LOCUS       CR605851                2748 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DE012YJ24 of Placenta of Homo sapiens
            (human).
ACCESSION   CR605851
VERSION     CR605851.1  GI:50486658
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2748)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2748)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2748
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YJ24"
                     /tissue_type="Placenta"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             90.1%;  Score 1333.8;  DB 3;  Length 2748;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||||| |||||| |||||| |||||||||
Db        772 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 831

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
```

```
                ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db        832 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 891

Qy        121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
              | || || ||| |||||||||||||||||||||| ||||||||| | |||||||||| | ||
Db        892 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 951

Qy        181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        952 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1011

Qy        241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1012 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1071

Qy        301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
              |||| |||||||||||||||||||||||||||| ||  ||||||||||||||||||||| ||
Db       1072 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1131

Qy        361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
              |||| ||||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db       1132 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1191

Qy        421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
               ||||| |||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1192 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1251

Qy        481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
              |||||||||||||| ||||||||||||||||||||||||| |||||||||||||||| |||
Db       1252 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1311

Qy        541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
              ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db       1312 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1371

Qy        601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1372 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1431

Qy        661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1432 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1491

Qy        721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
              ||||||||||| | ||||||||| |||||||||||||||||||||||||| |||||||||| |
Db       1492 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1551

Qy        781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1552 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1611

Qy        841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1612 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1671

Qy        901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
              ||||||||||||||||||| | ||||| || ||||||||| ||||||||||||||||||||
Db       1672 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1731

Qy        961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
              ||||||||||||||  ||||||||||||||||   |||||||||||||||||||| |||||
Db       1732 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1791

Qy       1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
              ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db       1792 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1851

Qy       1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
              ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1852 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1911

Qy       1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       1912 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1971

Qy       1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
              || ||||||  |||||||||||||||||||||||| |||| |||||||||||||||||||||
Db       1972 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2031

Qy       1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
              |||||||||||||||||| |||||||||||||||||| ||||||| | || ||||||||||
Db       2032 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2091

Qy       1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
              |||||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db       2092 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2151

Qy       1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||| 
Db       2152 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2211

Qy       1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
              ||||||||||| ||||| ||||      | ||| |  || |
Db       2212 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2252
```

```
RESULT 4
CR625046
LOCUS       CR625046                2748 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI044YK06 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR625046
VERSION     CR625046.1  GI:50505853
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2748)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2748)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2748
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI044YK06"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             90.1%;  Score 1333.8;  DB 3;  Length 2748;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
              ||||||||||||||||||| |||||||||||||||||||| ||||| |||||| ||||||||
Db        783 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 842

Qy         61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
              |||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         843 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 902

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db         903 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 962

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         963 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1022

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        1023 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1082

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        1083 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1142

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        1143 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1202

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        1203 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1262

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        1263 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1322

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        1323 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1382

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1383 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1442

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1443 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1502

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db        1503 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1562

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        1563 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1622

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1623 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1682

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db        1683 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1742

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        1743 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1802

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1803 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1862

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1863 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1922

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        1923 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1982

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db        1983 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2042

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db        2043 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2102

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        2103 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2162

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db        2163 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2222

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||| ||||| ||||      | ||| |  || |
Db        2223 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2263


RESULT 5
CR617248
LOCUS       CR617248                2749 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI022YJ18 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR617248
VERSION     CR617248.1  GI:50498055
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2749)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2749)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .2749
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI022YJ18"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             90.1%;  Score 1333.8;  DB 3;  Length 2749;
  Best Local Similarity   93.8%;  Pred. No. 0;
  Matches 1389;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db         777 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 836

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         837 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 896
```

```
Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db         897 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 956

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         957 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 1016

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        1017 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 1076

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db        1077 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 1136

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db        1137 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 1196

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        1197 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 1256

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db        1257 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 1316

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||| |||||||||||||  ||||||||
Db        1317 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATC 1376

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1377 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 1436

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1437 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 1496

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |
Db        1497 AGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAA 1556

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        1557 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTG 1616

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1617 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTAC 1676

Qy         901 ACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATT 960
               ||||||||||||||||||| | ||||| || ||||||||| |||||||||||||||||||
Db        1677 ACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATT 1736

Qy         961 CTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATC 1020
               ||||||||||||||  ||||||||||||||||   ||||||||||||||||||| |||||
Db        1737 CTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATC 1796

Qy        1021 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
               ||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |
Db        1797 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGG 1856

Qy        1081 GTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
               ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1857 GTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1916

Qy        1141 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 1200
               ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        1917 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTA 1976

Qy        1201 GGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAA 1260
               || |||||| ||||||||||||||||||||||| |||| |||||||||||||||||||||
Db        1977 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2036

Qy        1261 ATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
               |||||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||
Db        2037 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTC 2096

Qy        1321 CTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
               ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db        2097 CTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTG 2156

Qy        1381 TTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATT 1440
                |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| 
Db        2157 CTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATC 2216

Qy        1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
               ||||||||||| ||||| ||||      | ||| |  || |
Db        2217 GAAGCTGTAAAACTACAAATGGAGCCCAAGATGCAGTCCAA 2257


RESULT 6
BX391741/c
LOCUS       BX391741                 935 bp    mRNA    linear   EST 29-APR-2004
DEFINITION  BX391741 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YM13 3-PRIME, mRNA sequence.
ACCESSION   BX391741
VERSION     BX391741.1  GI:30619529
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAI022ZA10_CS02089_1&c=4215.r

FEATURES             Location/Qualifiers
     source          1..935
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             56.7%;  Score 839;  DB 5;  Length 935;
  Best Local Similarity   94.2%;  Pred. No. 5.6e-237;
  Matches  880;  Conservative    0;  Mismatches   53;  Indels    1;  Gaps    1;

Qy         225 TGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAG 284
               |||| ||||||||||||||||||||||||| ||||||||||    |||||||||||||||
Db         933 TGCCCCTCTTTGCATGCATGCAAATACTCA-TATTGGACAGAANNAATGATTAATCCTAG 875

Qy         285 TTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTC 344
```

```
                 |||||||| |||||||| || ||||||||||||||||||||||||||| ||  |||||||
Db          874 TTGTCCTGNAGGACTTGAAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTC 815

Qy          345 TGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCA 404
                ||||||||||||| |||||| |||||||||| ||||||| ||||| |||| | |||||||
Db          814 TGATGGGGGTGGAGTTCAAGATCAGGCAAGAAAAAAACATGTAAAAGAAGTATTCTCCCA 755

Qy          405 ACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACA 464
                ||| ||||||| |||| ||||| ||||||||||||||||||||| |||||||||||||||
Db          754 ACTCACCCGGGTACATGGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACA 695

Qy          465 TGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCT 524
                |||||||||||||||||||||||||||||| |||||||||||||||||||||||| ||||
Db          694 TGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCT 635

Qy          525 CCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAG 584
                |||||||||||| |||||||||||||||||||||||||| |||||||||||| |||||||
Db          634 CCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAG 575

Qy          585 GCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACAC 644
                ||||||  ||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db          574 GCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACAC 515

Qy          645 CACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCT 704
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          514 CACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCT 455

Qy          705 CACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGT 764
                |||||||||||||||||||||||||| | ||||||||| |||||||||||||||||||||
Db          454 CACCTGTGTAAAATTTAGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGT 395

Qy          765 AACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTC 824
                ||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db          394 AACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTC 335

Qy          825 AGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCC 884
                |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db          334 AGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCC 275

Qy          885 CCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAA 944
                ||||||||||||||||||||||||||||||||||| | ||||| || ||||||||| |||
Db          274 CCCTATGACCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAA 215

Qy          945 CAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTAC 1004
                ||||||||||||||||||||||||||||||  ||||||||||||||||   |||||||||
Db          214 CAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTAC 155

Qy         1005 TGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAA 1064
                |||||||||| |||||||||||||||||||||||||||||||||||||||||||| ||||
Db          154 TGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAA 95

Qy         1065 TGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCT 1124
                ||| |||||||||| |||| | ||||||||||||||||||||||||||||||||||||||
Db           94 TGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCT 35

Qy         1125 AGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTA 1158
                ||||||||||||||||||||||||||||||||||
Db           34 AGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTA 1


RESULT 7
BX347952
LOCUS       BX347952                 955 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX347952 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YM13 5-PRIME, mRNA sequence.
ACCESSION   BX347952
VERSION     BX347952.1  GI:30375235
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 955)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAE023ZF07_AE02167_1&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1. .955
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             54.9%;  Score 813.6;  DB 5;  Length 955;
  Best Local Similarity   93.7%;  Pred. No. 1.9e-229;
  Matches  890;  Conservative    0;  Mismatches   56;  Indels    4;  Gaps    4;

Qy          226 GCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGT 285
                ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db            1 GCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGT 60

Qy          286 TGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCT 345
                ||||||||||||||||||| ||||||||||||||||||||||||||| ||  ||||||||
Db           61 TGTCCTGGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCT 120

Qy          346 GATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAA 405
                |||||||||||| |||||| |||||||||| ||||||| ||||| |||| ||||||||||
Db          121 GATGGGGGTGGAGTTCAAGATCAGGCAAGAAAAAAACATGTAAAAGAAGTAATCTCCCAA 180

Qy          406 CTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACAT 465
                || ||||||| |||| ||||| ||||||||||||||||||||| ||||||||||||||||
Db          181 CTCACCCGGGTACATGGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACAT 240

Qy          466 GAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTC 525
                ||||||||||||||||||||||||||||| |||||||||||||||||||||||| |||||
Db          241 GAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTC 300

Qy          526 CATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGG 585
                ||||||||||| |||||||||||||||||||||||||| |||||||||||| ||||||||
Db          301 CATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGG 360

Qy          586 CCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACC 645
                |||||  ||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db          361 CCATATGTTTCAATCCCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACC 420

Qy          646 ACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTC 705
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 ACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTC 480

Qy          706 ACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTA 765
                ||||||||||||||||||||||||| | ||||||||| ||||||||||||||||||||||
Db          481 ACCTGTGTAAAATTTAGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTA 540

Qy          766 ACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCA 825
                || ||| ||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db          541 ACTCCT-CCAAACAAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCA 599
```

```
Qy      826 GCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCC 885
            ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      600 GCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCC 659

Qy      886 CCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAAC 945
            |||||||||||||||||||||||||||||||||| | ||||| || ||||||||| ||||
Db      660 CCTATGACCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAAC 719

Qy      946 AAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACT 1005
            |||||||||||||||||||||||||||||  ||||||||||||||||   ||||||||||
Db      720 AAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACT 779

Qy     1006 GGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAAT 1065
            ||||||||| |||||||||||||||||||||||||||||| ||||||||||||| |||||
Db      780 GGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACANACTATCTCAAGAACTAAAT 839

Qy     1066 GGTGACAT-GGAACAGGTCACTGACTCCCTGG-TCACCTTGCAAGATCAACTTAACTCCC 1123
            || ||||| ||||| | || | |||| ||||| |||||||||||||||||||||||  ||
Db      840 GGGGACATGGGAACGGTTCGCCGACTTCCTGGTTCACCTTGCAAGATCAACTTAACCTCC 899

Qy     1124 TAGCAGCAGTAG-TCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGC 1172
            ||||||||| || ||||||||||||||||||  |||||||||||| ||||
Db      900 TAGCAGCAGNAGTTCCTTCAAAATCGAAGAGGCTTAGACTTGCTACCCGC 949
```

```
RESULT 8
BX389656
LOCUS       BX389656                 931 bp    mRNA    linear   EST 29-APR-2004
DEFINITION  BX389656 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YM13 5-PRIME, mRNA sequence.
ACCESSION   BX389656
VERSION     BX389656.1  GI:30462930
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 931)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAF027ZB05_AF02531_1&c=4215.r
FEATURES             Location/Qualifiers
     source          1..931
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             50.8%;  Score 752.2;  DB 5;  Length 931;
  Best Local Similarity   92.3%;  Pred. No. 3.1e-211;
  Matches  856;  Conservative    0;  Mismatches   65;  Indels    6;  Gaps    6;
```

```
Qy      267 GAAAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTT 326
            ||||| |||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        1 GAAAAAGATTAATCCTAGTTGTCCTGGAGGACTTGGAGTCACTGTCTGTTGGACTTACTT 60

Qy      327 CACCCATACCAGTATGTCTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGT 386
            |||||| ||  |||||||||||||||||||| |||||| |||||||||| ||||||| ||
Db       61 CACCCAAACTGGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAAGAAAAAAACATGT 120

Qy      387 AAAGGAAGCAATCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACT 446
            ||| |||| |||||||||||| ||||||| |||| ||||| |||||||||||||||||||
Db      121 AAAAGAAGTAATCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCCTACAAAGGACT 180

Qy      447 AGTTCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAA 506
            || ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      181 AGATCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAA 240

Qy      507 TACCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTG 566
            ||||||||||||| |||||||||||||||| |||||||||||||||||||||||||| ||
Db      241 TACCACCCTCACTGGGCTCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATATG 300

Qy      567 CCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTT 626
            |||||||||| |||||||||||||  ||||||||||||| ||||||||||||||||||||
Db      301 CCTCCCCCTGAACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAACTT 360

Qy      627 CAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAAT 686
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAAT 420

Qy      687 AACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTC 746
            |||||||||||||||||||||||||||||||||||||||||||| | ||||||||| |||
Db      421 AACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAACTC 480

Qy      747 CCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATT 806
            ||||||||||||||||||||| |||||||||| |||||||||||||||||||||||||||
Db      481 CCAATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAATATT 540

Qy      807 TTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTT 866
            |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db      541 TTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTT 600

Qy      867 CCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGT 926
            ||||||||||||||||||||||||||||||||||||||||||||||||||||| | ||||
Db      601 CCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAGTTATGT 660

Qy      927 CGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGT 986
            | || ||||||||| ||||||||||||||||||||||||||||| |||  ||||||||||
Db      661 CATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGNTATAGGAGCAGGAGT 720

Qy      987 GCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAA 1046
            ||||||   ||||||||||||||||||| | ||||||||||||||||||||||||||||
Db      721 GCTAGGTGCACTAGGTACTGGCATTGGC-GGATCACAACCTCTACTCAGTTCTACTACAN 779

Qy     1047 ACTATCTCAAGAAATAAATGGTGAC-ATGGAACAGGTCACTGACT-CCCTGGTCACCTTG 1104
            ||||||||||||| ||||||| ||| ||| ||| |||| | |||| |||||||||||||
Db      780 ACTATCTCAAGAACTAAATGGGGACAATGTAACGGGTCGCCGACTCCCCTGGTCACCTTC 839

Qy     1105 CAAGATCAACTTAAC-TCCCTAGCAGCAGTA-GTCCTTCAAAATCGAAG-AGCTTTAGAC 1161
              ||||||||||| | ||||||||||||| | ||||||||||||||||| |||||||| |
Db      840 GCAGATCAACTTACCTTCCCTAGCAGCAGGAGGTCCTTCAAAATCGAAGAAGCTTTAGCC 899

Qy     1162 TTGCTAACCGCCAAAAGAGGGGGAACC 1188
            |||||||||  |   |   |||| | |
Db      900 TTGCTAACCCGCTGCAAGAGGGGGAAC 926
```

```
RESULT 9
BX326647/c
LOCUS       BX326647                 872 bp    mRNA    linear   EST 07-APR-2004
DEFINITION  BX326647 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YM13 3-PRIME, mRNA sequence.
ACCESSION   BX326647
VERSION     BX326647.1  GI:30334578
KEYWORDS    EST.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAD010ZC11_AD00948_1&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1..872
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             50.5%;  Score 747.2;  DB 5;  Length 872;
  Best Local Similarity   92.3%;  Pred. No. 9.3e-210;
  Matches  808;  Conservative    0;  Mismatches   64;  Indels    3;  Gaps    2;

Qy        278 ATCCTAGTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCA 337
              ||||||||||||||| |||||||| || ||||||||||||||||||||||  ||  ||
Db        872 ATCCTAGTTGTCCTGAAGGACTTGAAGTCACTGTCTGTTGGACTTACTTCCACCCAACTT 813

Qy        338 GTATGTCTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAA 397
              |        ||||||||||| |||||| |||||||||| ||||||| ||||| |||| ||
Db        812 GGTATGGCTATGGGGGTGGAGTTCAAGATCAGGCAAGAAAAAAACATGTAAA-GAAGTAA 754

Qy        398 TCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAA 457
              |||||||||| ||||||| |||| ||||| ||||||||||||||||||||| |||||||
Db        753 TCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAN 694

Qy        458 AACTACATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCA 517
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        693 AACTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATACCACCCTCA 634

Qy        518 CTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGC 577
              || |||||||||||||||| |||||||||||||||||||||||||| ||||||||||||
Db        633 CTGGGCTCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGA 574

Qy        578 ACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAA 637
              |||||||||||||  ||||||||||||| |||||||||||||||||||||||||||||||
Db        573 ACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAACTTCAGCACAGAAA 514

Qy        638 TAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCT 697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 TAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCT 454

Qy        698 CAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCA 757
              ||||||||||||||||||||||||||||||||| | ||||||||| ||||||||||||||
Db        453 CAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAACTCCCAATGCATCA 394

Qy        758 GGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTG 817
              |||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||
Db        393 GGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTG 334

Qy        818 GTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCT 877
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        333 GTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCT 274

Qy        878 TAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGC 937
              |||||||||||||||||||||||||||||||||||||||||| | ||||| || ||||||
Db        273 TAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAAGC 214

Qy        938 CCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGAC 997
              ||| |||||||||||||||||||||||||||||||||  |||||||||||||||||   ||
Db        213 CCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCAC 154

Qy        998 TAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAG 1057
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        153 TAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAG 94

Qy       1058 AAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTA 1117
              || ||||||| |||||||||| |||| | ||||||   |    | ||||||||||||||||
Db         93 AACTAAATGGGGACATGGAACGGGTCGCCGACTCC--TGGTCACCTGCAAGATCAACTTA 36

Qy       1118 ACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGA 1152
              |||||||||||||||||||||||||||||||||||
Db         35 ACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGA 1


RESULT 10
BX408752
LOCUS       BX408752                 903 bp    mRNA    linear   EST 01-MAY-2004
DEFINITION  BX408752 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE013YI20
            5-PRIME, mRNA sequence.
ACCESSION   BX408752
VERSION     BX408752.2  GI:46931100
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 13, 2003 this sequence version replaced gi:30648825.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAA008ZH03_CS00697_2&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1..903
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE013YI20"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             50.0%;  Score 741;  DB 5;  Length 903;
  Best Local Similarity   93.0%;  Pred. No. 6.5e-208;
  Matches  796;  Conservative    0;  Mismatches   58;  Indels    2;  Gaps    2;

Qy          4 GCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACTGCA 63
              | ||  || |   ||| |||||||||||||||||| |||||| ||||| |||||||||||
Db         50 GGCCCTCCCTTANATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACTGCA 109

Qy         64 CCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGAACG 123
              |||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||| |
Db        110 CCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGAATG 169

Qy        124 CGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCCACC 183
              | || ||| |||||||||||||||||||| ||||||||| | |||||||||| | |||||
Db        170 CAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCCACC 229

Qy        184 TTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATGCAT 243
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        230 TTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATGCAT 289

Qy        244 GCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTTGGA 303
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        290 GCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTTGGA 349

Qy        304 GCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATTCAA 363
              | ||||||||||||||||||||||||||| ||  |||||||||||||||||||| |||||
Db        350 GTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTTCAA 409

Qy        364 GGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACATAGC 423
              | |||||||||||||||||| ||||| |||| |||||||||||| ||||||| |||| ||
Db        410 GATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACATGGC 469

Qy        424 ACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACCCAT 483
              ||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        470 ACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACCCAT 529

Qy        484 ACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAA 543
              ||||||||||| |||||||||||||||||||||||| |||||||||||||||| ||||||
Db        530 ACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCCCAA 589

Qy        544 AACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCT 603
              |||||||||||||||||||| ||||| |||||| |||||||||||||  |||||||||||
Db        590 AACCCTACTAACTGTTGGATATGCCTNCCCCTGAACTTCAGGCCATATGTTTCAATCCCT 649

Qy        604 GTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGA 663
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        650 GTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGA 709

Qy        664 CCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGC 723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        710 CCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGC 769

Qy        724 AATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATA 783
              ||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| ||||
Db        770 AATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAAATA 829

Qy        784 GTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAAT 843
              ||||||||||||||||| |||||||||||||| |||||||||||||||   |||||| ||
Db        830 GTCTGCCTACCCTCAGG-ATATTTTTTGTCTG-GGTACCTCAGCCTATTCGTGTTTGGAT 887

Qy        844 GGCTCTTCAGAATCTA 859
              ||||||||||||||||
Db        888 GGCTCTTCAGAATCTA 903


RESULT 11
BX408733
LOCUS       BX408733                 883 bp    mRNA    linear   EST 30-APR-2004
DEFINITION  BX408733 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012YJ24
            5-PRIME, mRNA sequence.
ACCESSION   BX408733
VERSION     BX408733.1  GI:30635957
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAA008ZE02_CS00686_1&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1..883
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YJ24"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             49.8%;  Score 737.2;  DB 5;  Length 883;
  Best Local Similarity   93.2%;  Pred. No. 8.7e-207;
  Matches  825;  Conservative    0;  Mismatches   54;  Indels    6;  Gaps    5;

Qy         29 TTACTGTTCTCTTACCCCCTTTCGCTCTCACTGCACCCCCTCCATGCTGCTGTACAACCA 88
              |||||||||| |||||| ||||       |||||||||||||||||| ||||||  ||||
Db          1 TTACTGTTCTTTTACCCTCTTT--ACTCTACTGCACCCCCTCCATGCCGCTGTATGACCA 58

Qy         89 GTAGCTCCCCTTACCAAGAGTTTCTATGAAGAACGCGGCTTCCTGGAAATATTGATGCCC 148
              |||||||||||||||||||||||||||| |||| || || ||| ||||||||||||||||
Db         59 GTAGCTCCCCTTACCAAGAGTTTCTATGGAGAATGCAGCGTCCCGGAAATATTGATGCCC 118

Qy        149 CATCATATAGGAGTTTATCTAAGGGAAACTCCACCTTCACTGCCCACACCCATATGCCCC 208
              |||| ||||||||| | |||||||||| | ||||||||||||||||||||||||||||||
Db        119 CATCGTATAGGAGTCTTTCTAAGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCC 178

Qy        209 GCAACTGCTATAACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGA 268
              ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |
Db        179 GCAACTGCTATCACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAA 238

Qy        269 AAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCA 328
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        239 AAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTCA 298

Qy        329 CCCATACCAGTATGTCTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAA 388
              |||| ||  |||||||||||||||||||| |||||| |||||||||||||||||| ||||
Db        299 CCCAAACTGGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAAGAGAAAAACATGTAA 358

Qy        389 AGGAAGCAATCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAG 448
              | |||| |||||||||||| ||||||| |||| ||||| |||||||||||||||||||||
Db        359 AAGAAGTAATCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCCTACAAAGGACTAG 418

Qy        449 TTCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATA 508
               ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db        419 ATCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATA 478

Qy        509 CCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCC 568
              ||||||||||| |||||||||||||||| |||||||||||||||||||||||||| ||||
Db          479 CCACCCTCACTGGGCTCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATATGCC 538

Qy          569 TCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCA 628
                |||||||| |||||||||||||  ||||||||||||| ||||||||||||||||||||||
Db          539 TCCCCCTGAACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAACTTCA 598

Qy          629 GCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAA 688
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          599 GCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAA 658

Qy          689 CCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCC 748
                |||||||||||||||||||||||||||||||||||||||||| | ||||||||| || ||
Db          659 CCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAACTNCC 718

Qy          749 AATGCATCA-GGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTT 807
                ||||||||| |||||||||| ||| |||||| ||||||||||||||||||| ||||||||
Db          719 AATGCATCAGGGTGGGTAACTCCTTCCACACAAATAGTCTGCCTACCCTCA-GAATATTT 777

Qy          808 TTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTC-TTCAGAATCTATGTGCTT 866
                ||||||||||||||||||||||||| ||||||||||||||| ||||||||||||||||||
Db          778 TTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTTCAGAATCTATGTGCTT 837

Qy          867 CCTCTCATTCTTAGTGCCCCC-TATGACCATCTACACTGAACAAG 910
                ||||||||||||||||||||| ||||||  |||||||||||||||
Db          838 CCTCTCATTCTTAGTGCCCCCTTATGACATTCTACACTGAACAAG 882

RESULT 12
BX347619/c
LOCUS       BX347619                 877 bp    mRNA    linear   EST 23-APR-2004
DEFINITION  BX347619 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI022YJ18 3-PRIME, mRNA sequence.
ACCESSION   BX347619
VERSION     BX347619.1  GI:30379218
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 877)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAD009ZB06_AD00811_1&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1..877
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI022YJ18"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             49.2%;  Score 728.4;  DB 5;  Length 877;
  Best Local Similarity   93.2%;  Pred. No. 3.5e-204;
  Matches  803;  Conservative    0;  Mismatches   55;  Indels    4;  Gaps    4;
```

```
Qy          503 TTAATACCACCCTCACTCGGCTCCATGAGGTCTC-AGCCCAAAACCCTACTAACTGTTGG 561
                | ||  | ||||||||| |||||||||||  |||  ||| ||||||||||||||||| ||
Db          877 TAAAACCAACCCTCACTGGGCTCCATGAGTNCTCGGGCCAAAAACCCTACTAACTGTGGG 818

Qy          562 ATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAAC 621
                || ||||||| |||| |||||||| ||||  ||||||||||||| |||||||||||||||
Db          817 ATATGCCTCCNCCTGAACTTCAGGGCATATGTTTCAATCCCTGTACCTGAACAATGGAAC 758

Qy          622 AACTTCAGCACAGA-AATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCT 680
                |  ||||||||||| ||||||||||||| |||||||||||||||||||||||||||||||
Db          757 ACTTTCAGCACAGANAATAAACACCACTNCCGTTTTAGTAGGACCTCTTGTTTCCAATCT 698

Qy          681 GGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAAC 740
                ||||||||||||||||||||||||||||||||| |||||||||||||||| | |||||||
Db          697 GGAAATAACCCATACCTCAAACCTCACCTGTGT-AAATTTAGCAATACTACATACACAAC 639

Qy          741 CAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGG 800
                || ||| |||||||||||||||||||| |||||||||| |||||||||||||||||||||
Db          638 CAACTCNCAATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGG 579

Qy          801 AATATTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTAT 860
                ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db          578 AATATTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTAT 519

Qy          861 GTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAA 920
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          518 GTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAG 459

Qy          921 TCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGC 980
                | ||||| || ||||||||| |||||||||||||||||||||||||||||||||  ||||
Db          458 TTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGC 399

Qy          981 AGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTA 1040
                ||||||||||||   ||||||||||||||||||| |||||||||||||||||||||||||
Db          398 AGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTA 339

Qy         1041 CTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCAC 1100
                ||||||||||||||||||| ||||||| |||||||||| |||| | ||||||||||||||
Db          338 CTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCAC 279

Qy         1101 CTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGA 1160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          278 CTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGA 219

Qy         1161 CTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTA 1220
                ||||||||||||  |||||||||||||||||||||||||||| |||||| ||||||||||
Db          218 CTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTAGGGGAAGAATGCTGTTATTA 159

Qy         1221 TGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATG 1280
                ||||||||||||| |||| ||||||||||||||||||||||||||||||||||||||| |
Db          158 TGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAACG 99

Qy         1281 TAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTG 1340
                ||||||||||||||||| |||||| | || |||||||||||||||||||||||||||| |
Db           98 TAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCC-G 40

Qy         1341 GGTTCTCCCCTTCTTAGGACCT 1362
                | ||||||||||||||||||||
Db           39 GATTCTCCCCTTCTTAGGACCT 18

RESULT 13
BX388766/c
LOCUS       BX388766                 836 bp    mRNA    linear   EST 29-APR-2004
DEFINITION  BX388766 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI044YK06 3-PRIME, mRNA sequence.
ACCESSION   BX388766
VERSION     BX388766.2  GI:46877917
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 836)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 8, 2003 this sequence version replaced gi:30462383.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AU012ZB04_U01082_1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..836
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI044YK06"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             48.2%;  Score 714;  DB 5;  Length 836;
  Best Local Similarity   92.6%;  Pred. No. 6.5e-200;
  Matches  772;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    2;

Qy        565 TGCCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAAC 624
              |||||||||||| |||  |||||| |  || ||||||||| |||| | ||||||  || |
Db        835 TGCCTCCCCCTGAACTCAAGGCCAAATGTTCCAATCCCTGGTCCTCACCAATGGCCCACC 776

Qy        625 TTCAGCACAGAAA-TAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGA 683
              |||||| |||||| || |||||||| |||||| |||||||||||||||||||||||||||
Db        775 TTCAGCCCAGAAATTACACACCACTCCCGTTTAAGTAGGACCTCTTGTTTCCAATCTGGA 716

Qy        684 AATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAG 743
              |||||||||| ||||||||||||||||||||||||| |||||| ||| | ||||| |||
Db        715 AATAACCCATTCCTCAAACCTCACCTGTGTAAAATTAAGCAATCCTACATACACACCCAA 656

Qy        744 CTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAAT 803
              |||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db        655 CTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAAT 596

Qy        804 ATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTG 863
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        595 ATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTG 536

Qy        864 CTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCA 923
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db        535 CTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAGTTA 476

Qy        924 TGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGG 983
              |||| || ||||||||| |||||||||||||||||||||||||||||||||  |||||||
Db        475 TGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGG 416

Qy        984 AGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTA 1043
              |||||||||   ||||||||||||||||||| ||||||||||||||||||||||||||||
Db        415 AGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTA 356

Qy       1044 CAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTT 1103
              |||||||||||||||| ||||||| |||||||||| |||| | |||||||||||||||||
Db        355 CAAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTT 296

Qy       1104 GCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTT 1163
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 GCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTT 236

Qy       1164 GCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGT 1223
              |||||||||  |||||||||||||||||||||||||||| |||||| |||||||||||||
Db        235 GCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTAGGGGAAGAATGCTGTTATTATGT 176

Qy       1224 TAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAG 1283
              |||||||||| |||| ||||||||||||||||||||||||||||||||||||||| ||||
Db        175 TAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAACGTAG 116

Qy       1284 AGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGT 1343
              |||||||||||||| |||||| | || ||||||||||||||||||||||||||||||| |
Db        115 AGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGAT 56

Qy       1344 TCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGACC 1397
              |||||||||||||||||||||||||||| ||||| || |||||| ||| || ||
Db         55 TCTCCCCTTCTTAGGACCTCTAGCAGCTATAATA-TGCTACTCCACTTGGGCCC 3


RESULT 14
BX408734
LOCUS       BX408734                 921 bp    mRNA    linear   EST 30-APR-2004
DEFINITION  BX408734 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012YJ24
            5-PRIME, mRNA sequence.
ACCESSION   BX408734
VERSION     BX408734.1  GI:30635959
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAA008ZE02_CS00686_2&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1..921
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YJ24"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             47.9%;  Score 709.8;  DB 5;  Length 921;
  Best Local Similarity   88.6%;  Pred. No. 1.2e-198;
  Matches  815;  Conservative    0;  Mismatches   98;  Indels    7;  Gaps    4;

Qy         29 TTACTGTTCTCTTACCCCCTTTCGCTCTCACTGCACCCCCTCCATGCTGCTGTACAACCA 88
              || ||||||| |||||| ||||| ||||||||||||||||||||||| ||||||  ||||
Db          1 TTTCTGTTCTTTTACCCTCTTTCACTCTCACTGCACCCCCTCCATGCCGCTGTATGACCA 60

Qy         89 GTAGCTCCCCTTACCAAGAGTTTCTATGAAGAACGCGGCTTCCTGGAAATATTGATGCCC 148
```

```
              ||||||||||||||||| |||||||||| |||| || || ||| || |||||||||||||
Db         61 GTAGCTCCCCTTACCAANAGTTTCTATGGAGAATGCAGCGTCCCGG-AATATTGATGCCC 119

Qy        149 CATCATATAGGAGTTTATCTAAGGGAAACTCCACCTTCACTGCCCACACCCATATGCCCC 208
              |||| ||||||||| | |||||||||| | ||||||||||||||||||||||||||||||
Db        120 CATCGTATAGGAGTCTTTCTAAGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCC 179

Qy        209 GCAACTGCTATAACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGA 268
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||| |
Db        180 GCAACTGCTATCACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAA 239

Qy        269 AAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCA 328
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        240 AAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTCA 299

Qy        329 CCCATACCAGTATGTCTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAA 388
              |||| ||  |||||||||||||||||||| |||||| |||||||||||||||||| ||||
Db        300 CCCAAACTGGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAAGAGAAAAACATGTAA 359

Qy        389 AGGAAGCAATCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAG 448
              | |||| |||||||||||| ||||||| |||| ||||| |||||||||||||||||||||
Db        360 AAGAAGTAATCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCCTACAAAGGACTAG 419

Qy        449 TTCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATA 508
               ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db        420 ATCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATA 479

Qy        509 CCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCC 568
              ||||||||||| |||||||||||||||| |||||||||||||||||||||||||| ||||
Db        480 CCACCCTCACTGGGCTCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATATGCC 539

Qy        569 TCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCA 628
              |||||||| ||||||||||||||  ||||||||||||| ||||||||||||||||||||||
Db        540 TCCCCCTGAACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAACTTCA 599

Qy        629 GCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAA 688
              |||||||||| ||||||||||||  |||||||||||||||||||||||||||||||||||
Db        600 GCACAGAAATTAACACCACTTCCCGTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAA 659

Qy        689 CCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCC 748
              |||||||||||||||||||||||||||||||||||||||||| | ||||||||| || ||
Db        660 CCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCA-CTTCC 718

Qy        749 AATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTT 808
              |||| || ||||||||||| ||| || ||| ||||||||||||||||||| | |||||||
Db        719 AATGGATTAGGTGGGTAACTCCTTCCCCACAAATAGTCTGCCTACCCTCAAGGATATTTT 778

Qy        809 TTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCC 868
              ||| |||||| |||||||||||||   | ||| |||| |||||||||  || | ||||||
Db        779 TTGGCTGTGGGACCTCAGCCTATCGGCGGTTGGATGGGTCTTCAGAA-ATAAGGGCTTCC 837

Qy        869 TCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCG 928
               ||||| ||  |  |||| |    || ||        | |    |||||||||    ||
Db        838 CCTCATCCTAGGGCCCCCATGACCCCTTCACTGGAACAGA----TTATACAATTTGTTCT 893

Qy        929 TACCTAAGCCCCACAACAAA 948
                 |||| |||||| |||||
Db        894 ATACTAACCCCCACCACAAA 913


RESULT 15
BX347111/c
LOCUS       BX347111                 844 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX347111 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI022YE21 3-PRIME, mRNA sequence.
ACCESSION   BX347111
VERSION     BX347111.1  GI:30375145
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            7108.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AU010ZB08_U0922_1&c=7108.r.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI022YE21"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             47.7%;  Score 705.8;  DB 5;  Length 844;
  Best Local Similarity   92.7%;  Pred. No. 1.8e-197;
  Matches  762;  Conservative    0;  Mismatches   58;  Indels    2;  Gaps    2;

Qy        502 TTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAA-CCCTACTAACTGTTG 560
              || ||| |||||||| || || ||||| ||||||| |||||||| ||||||||||||| |
Db        837 TTAAATCCCACCCTCCCTGGGTTCCATTAGGTCTCGGCCCAAAACCCCTACTAACTGTGG 778

Qy        561 GATGTGCCT-CCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGA 619
              ||| ||||| ||||||| | |||||||||||  ||||||||||||| |||||||||||||
Db        777 GATATGCCTCCCCCCTGAATTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGA 718

Qy        620 ACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATC 679
                |  |||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        717 CAACTTTCAGCACAGAAATAAACCCCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATC 658

Qy        680 TGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAA 739
              ||||||||||||||||||||||||||||||||||||||||||||||||||| | ||||||
Db        657 TGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAA 598

Qy        740 CCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAG 799
              ||| ||| |||||||||||||||||||| |||||||||| ||||||||||||||||||||
Db        597 CCAACTCNCAATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAG 538

Qy        800 GAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTA 859
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        537 GAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTA 478

Qy        860 TGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACA 919
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        477 TGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACA 418

Qy        920 ATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAG 979
               | ||||| || ||||||||| |||||||||||||||||||||||||||||||||  |||
Db        417 GTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAG 358

Qy        980 CAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCT 1039
              |||||||||||||   ||||||||||||||||||| ||||||||||||||||||||||||
Db        357 CAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCT 298

Qy       1040 ACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCA 1099
              |||||||||||||||||||| ||||||| |||||||||| |||| | |||||||||||||
Db        297 ACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCA 238
```

```
Qy       1100 CCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAG 1159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 CCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAG 178

Qy       1160 ACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATT 1219
              |||||||||||||  |||||||||||||||||||||||||||| |||||| |||||||||
Db        177 ACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTAGGGGAAGAATGCTGTTATT 118

Qy       1220 ATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAAT 1279
              |||||||||||||| |||| |||||||||||||||||||||||||||||||||||||||
Db        117 ATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAAC 58

Qy       1280 GTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCC 1321
              |||||||||||||||||| |||||| | ||   ||| |  ||
Db         57 GTAGAGCAGAGGAGCTTCGAAACACTGGACCTGGGGCCGCCC 16


RESULT 16
BX367907
LOCUS       BX367907                 828 bp    mRNA    linear   EST 28-APR-2004
DEFINITION  BX367907 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YM13 5-PRIME, mRNA sequence.
ACCESSION   BX367907
VERSION     BX367907.1  GI:30459586
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AT022ZG01_T02023_1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..828
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             47.5%;  Score 703.2;  DB 5;  Length 828;
  Best Local Similarity   93.8%;  Pred. No. 1e-196;
  Matches  754;  Conservative    0;  Mismatches   48;  Indels    2;  Gaps    2;

Qy        110 TTCTATGAAGAACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTA 169
              ||   || |||| || || ||| || ||||||||||||||||| ||||||||| | ||||
Db          9 TTTCTTGGAGAATGCAGCGTCCCGG-AATATTGATGCCCCATCGTATAGGAGTCTTTCTA 67

Qy        170 AGGGAAACTCCACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCA 229
              |||||| | ||||||||||||||||||||||||||||||||||||||||| |||||||||
Db         68 AGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCA 127

Qy        230 CTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTC 289
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        128 CTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTC 187

Qy        290 CTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATG 349
              ||||||||||||||| ||||||||||||||||||||||||||| ||  ||||||||||||
Db        188 CTGGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATG 247

Qy        350 GGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGA 409
              |||||||| |||||| |||||||||| ||||||| ||||| |||| |||||||||||| |
Db        248 GGGGTGGAGTTCAAGATCAGGCAAGAAAAAAACATGTAAAAGAAGTAATCTCCCAACTCA 307

Qy        410 CCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAA 469
              |||||| |||| ||||| ||||||||||||||||||||| ||||||||||||||||||||
Db        308 CCCGGGTACATGGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAA 367

Qy        470 CCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATG 529
              ||||||||||||||||||||||||| |||||||||||||||||||||||| |||||||||
Db        368 CCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATG 427

Qy        530 AGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCAT 589
              ||||||| |||||||||||||||||||||||||| |||||||||||| ||||||||||||
Db        428 AGGTCTCGGCCCAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCAT 487

Qy        590 ACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTT 649
              |  ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        488 ATGTTTCAATCCCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTT 547

Qy        650 CCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCT 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 CCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCT 607

Qy        710 GTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACAC 769
              ||||||||||||||||||||| | ||||||||| |||||||||||||||||||||||| |
Db        608 GTGTAAAATTTAGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTC 667

Qy        770 CTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCT 829
              ||||||||| |||||||||||||||||||||||||||||||||||||| ||||||| |||
Db        668 CTCCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGTTACCTCAACCT 727

Qy        830 ATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTA 889
              ||| ||||||||||||||| |||||||||||| |||||||||||||||||| ||||||||
Db        728 ATCGTTGTTTGAATGGCTC-TCAGAATCTATGGGCTTCCTCTCATTCTTAGGGCCCCCTA 786

Qy        890 TGACCATCTACACTGAACAAGATT 913
              ||||||||||||||||||||||||
Db        787 TGACCATCTACACTGAACAAGATT 810


RESULT 17
BX368078/c
LOCUS       BX368078                 846 bp    mRNA    linear   EST 26-APR-2004
DEFINITION  BX368078 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YM13 3-PRIME, mRNA sequence.
ACCESSION   BX368078
VERSION     BX368078.2  GI:46572908
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 8, 2003 this sequence version replaced gi:30447710.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AU012ZD05_U01092_1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..846
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             47.2%;  Score 698.6;  DB 5;  Length 846;
  Best Local Similarity   92.8%;  Pred. No. 2.4e-195;
  Matches  786;  Conservative    0;  Mismatches   56;  Indels    5;  Gaps    5;

Qy        467 AAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATACCAC-CCTCACTCGGCTC 525
              |||||  | ||||||||||||| || || ||||||||||||| ||  ||||||| |||||
Db        846 AAACCTCCGGTACCCATACTCGGCTTGTAAGCCTATTTAATAACAAACCTCACTGGGCTC 787

Qy        526 CATGAGGTCTCAGCCC-AAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCA- 583
                  ||||||| |||| ||||||||||||||||| | || |||||||||||| ||||||
Db        786 ATATAGGTCTCGGCCCAAAAACCCTACTAACTGTGGAATATGCCTCCCCCTGAACTTCAG 727

Qy        584 GGCCATACATTTCAAT-CCCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAAC 642
              |||||||  ||||||| |||||| ||||| ||||||||||||||||||||||||||||||
Db        726 GGCCATATGTTTCAATNCCCTGTACCTGANCAATGGAACAACTTCAGCACAGAAATAAAC 667

Qy        643 ACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAAC 702
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 ACCACTNCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAAC 607

Qy        703 CTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGG 762
              |||||||||||||||||||||||||||| | ||||||||| |||||||||||||||||||
Db        606 CTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGG 547

Qy        763 GTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGTGGTACC 822
              ||||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        546 GTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGTGGTACC 487

Qy        823 TCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTG 882
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 TCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTG 427

Qy        883 CCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCAC 942
              ||||||||||||||||||||||||||||||||||||| | ||||| || ||||||||| |
Db        426 CCCCCTATGACCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGC 367

Qy        943 AACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGT 1002
              ||||||||||||||||||||||||||||||||  ||||||||||||||||   |||||||
Db        366 AACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGT 307

Qy       1003 ACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATA 1062
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||| ||
Db        306 ACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTA 247

Qy       1063 AATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCC 1122
              ||||| |||||||||| |||| | ||||||||||||||||||||||||||||||||||||
Db        246 AATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCC 187

Qy       1123 CTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGG 1182
              ||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db        186 CTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGG 127

Qy       1183 GGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTC 1242
              |||||||||||||||||||| |||||| ||||||||||||||||||||||| |||| |||
Db        126 GGAACCTGTTTATTTTTAGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTC 67

Qy       1243 ACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAAC 1302
              |||||||||||||||||||||||||||||||||||| || ||||||||||||||| ||||
Db         66 ACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAACGT-GAGCAGAGGAGCTTCGAAAC 8

Qy       1303 ACCGAAC 1309
              |  || |
Db          7 ATGGACC 1


RESULT 18
BX389657
LOCUS       BX389657                 995 bp    mRNA    linear   EST 29-APR-2004
DEFINITION  BX389657 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YM13 5-PRIME, mRNA sequence.
ACCESSION   BX389657
VERSION     BX389657.2  GI:46875146
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 995)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 8, 2003 this sequence version replaced gi:30462931.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAF027ZB05_AF02531_2&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1..995
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             47.0%;  Score 696.2;  DB 5;  Length 995;
  Best Local Similarity   86.8%;  Pred. No. 1.3e-194;
  Matches  826;  Conservative    0;  Mismatches  114;  Indels   12;  Gaps    5;

Qy        266 GGAAAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACT 325
              | |||||||||||||||||||||||||||||||||||||  |||||||||||||||||||
Db          1 GAAAAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGT-ACTGTCTGTTGGACTTACT 59

Qy        326 TCACCCATACCAGTATGTCTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAG 385
              ||||||| ||  |||||||||||||||||||| |||||| |||||||||| ||||||| |
Db         60 TCACCCAAACTGGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAAGAAAAAAACATG 119

Qy        386 TAAAGGAAGCAATCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGAC 445
              |||| |||| |||||||||||| ||||||| |||| ||||| ||||||||||||||||||
Db        120 TAAAAGAAGTAATCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCCTACAAAGGAC 179

Qy        446 TAGTTCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTA 505
              ||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        180 TAGATCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTA 239
```

```
Qy         506 ATACCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGT 565
               |||||||||||||| |||||||||||||||| |||||||||||||||||||||||||| |
Db         240 ATACCACCCTCACTGGGCTCCATGAGGTCTCGGCCCAAAACCCTACTAACTGTTGGATAT 299

Qy         566 GCCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACT 625
               ||||||||||| |||||||||||||  ||||||||||||| |||||||||||||||||||
Db         300 GCCTCCCCCTGAACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAACT 359

Qy         626 TCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAA 685
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         360 TCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAA 419

Qy         686 TAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCT 745
               ||||||||||||||||||||||||||||||||||||||||||||| | ||||||||| ||
Db         420 TAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAACT 479

Qy         746 CCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATAT 805
               |||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||||
Db         480 CCCAATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAATAT 539

Qy         806 TTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCT 865
               ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db         540 TTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCT 599

Qy         866 TCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATG 925
               | |    ||||||||||||||||||||||||||||||||||||||||||||||| | |||
Db         600 TGCCTCTATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAGTTATG 659

Qy         926 TCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAG 985
               || || ||||  ||| ||||||||||||||||||||||||||||  |||  ||    |||
Db         660 TCATATCTAAAGCCCGCAACAAAAGAGTACCCATTCTTCCTTTTTGTATAGGAAGCAGAG 719

Qy         986 TGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACA 1045
               |||||||   ||||| | ||| ||||||  | || ||| ||| || ||| ||||||||||
Db         720 TGCTAGGTGCACTAGATCCTGCCATTGGTGGAATAACATCCTTTATTCATTTCTACTACA 779

Qy        1046 AACTATCTCAA-GAAATAAATGGTGACATGGAACAGGT---CACTGACTCCCTGGTCACC 1101
               |||||| | || ||  |||||||  ||||||||| |||   |     | ||  |    |
Db         780 AACTATTTNAAGGACCTAAATGGGAACATGGAACGGGTTGGCCGACTCCCCTGGTAACCT 839

Qy        1102 TTGCAAGATCAACTTAACTCCC----TAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTT 1157
               | | |||||||||||||||  |     ||| | ||      |  |||    |||  ||||
Db         840 TGGAAAGATCAACTTAACTTGCCTAACAGCCGGAGGTTCTTTCAAAATCGCAAGGCCTTT 899

Qy        1158 AGACTTGCTAACCGC---CAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAA 1206
               |||||  ||||||||    || || |||| | | |||||||||||||| |||
Db         900 AGACTGTCTAACCGCTGTAAAGAGGGGGGAACCTTGTTTATTTTTAGGGGAA 951


RESULT 19
BX347314
LOCUS       BX347314                 924 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX347314 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI044YK06 5-PRIME, mRNA sequence.
ACCESSION   BX347314
VERSION     BX347314.1  GI:30365192
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAA016ZH06_CS01489_1&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI044YK06"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             46.9%;  Score 695;  DB 5;  Length 924;
  Best Local Similarity   92.7%;  Pred. No. 2.9e-194;
  Matches  750;  Conservative    0;  Mismatches   57;  Indels    2;  Gaps    2;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               ||||||||||||||||||| |||||||||||||||||| |||||| ||||| ||||||||
Db         112 ATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 171

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         172 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 231

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db         232 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 291

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         292 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 351

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db         352 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 411

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db         412 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 471

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db         472 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 531

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         532 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 591

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db         592 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 651

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               ||||||||||||||||||||||| |||||||||||| |||||| ||||||  ||||||||
Db         652 CAAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCA-GCCATATGTTTCAATC 710

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| ||||||||||||||||||||||||||||||||||||||||||| |||| |||||
Db         711 CCTGTACCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTNCGTTNTAGTA 770

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||| | |||||| |||||||||||||||||||||||||||||||||||||
Db         771 GGACCTCTTGTTTTCCATCTGGGAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 830

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               ||| |||||| | |||||||||  | || |||||||| |  |||||| |||||||||| |
Db       831 AGCCATACTACATACACAACCAAATTCCCATGCATCAAGGTGGTAACTCCTCCCACAC-A 889

Qy       781 ATAGTCTGCCTACCCTCAGGAATATTTTT 809
             |||||||||||||||||||||||||||||
Db       890 ATAGTCTGCCTACCCTCAGGAATATTTTT 918


RESULT 20
AU138405
LOCUS       AU138405                 723 bp    mRNA    linear   EST 02-AUG-2002
DEFINITION  AU138405 PLACE1 Homo sapiens cDNA clone PLACE1008489 5', mRNA
            sequence.
ACCESSION   AU138405
VERSION     AU138405.1  GI:10999926
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1008489"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="Vector: pME18SFL3"
ORIGIN

  Query Match             43.7%;  Score 647;  DB 1;  Length 723;
  Best Local Similarity   94.6%;  Pred. No. 4.7e-180;
  Matches  679;  Conservative    0;  Mismatches   38;  Indels    1;  Gaps    1;

Qy       565 TGCCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAAC 624
             |||||||||||| |||||||||||||  ||||||||||||| ||||||||||||||||||
Db         3 TGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAAC 62

Qy       625 TTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAA 684
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63 TTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAA 122

Qy       685 ATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGC 744
             |||||||||||||||||||||||||||||||||||||||||||||| | ||||||||| |
Db       123 ATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAAC 182

Qy       745 TCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATA 804
             ||||||||||||||||||||||| |||||||||| |||||||||||||||||||||||||
Db       183 TCCCAATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAATA 242

Qy       805 TTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGC 864
             |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       243 TTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGC 302

Qy       865 TTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCAT 924
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| | ||
Db       303 TTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAGTTAT 362

Qy       925 GTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGA 984
             ||| || ||||||||| |||||||||||||||||||||||||||||||||  ||||||||
Db       363 GTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGGA 422

Qy       985 GTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTAC 1044
             ||||||||   ||||||||||||||||||| |||||||||||||||||||||||||||||
Db       423 GTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTAC 482

Qy      1045 AAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTG 1104
             ||||||||||||||| ||||||| |||||||||| |||| | ||||||||||||||||||
Db       483 AAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTTG 542

Qy      1105 CAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTG 1164
             |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       543 CAAGATNAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTG 602

Qy      1165 CTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTT 1224
             ||||||||  ||||| |||||||||||||| ||||||| |||||| |||||||||| |||
Db       603 CTAACCGCTGAAAGANGGGGAACCTGTTTAATTTTAGGGGAAGAATGCTGTTATTANGTT 662

Qy      1225 AATCAATCCAGAATTGTCACT-GAGAAAGTTAAAGAAATTCGAGATCGAATACAATGT 1281
             ||||||||| | || |||||| ||||||||||||||||||||||||||||||||| ||
Db       663 AATCAATCCGGGATCGTCACTGGAGAAAGTTAAAGAAATTCGAGATCGAATACAACGT 720


RESULT 21
AQ261133
LOCUS       AQ261133                 631 bp    DNA     linear   GSS 24-OCT-1998
DEFINITION  CITBI-E1-2506C15.TF CITBI-E1 Homo sapiens genomic clone 2506C15,
            genomic survey sequence.
ACCESSION   AQ261133
VERSION     AQ261133.1  GI:3787657
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CITBI-E1-2506C15.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2506C15"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
ORIGIN

  Query Match             42.5%;  Score 629.4;  DB 8;  Length 631;
  Best Local Similarity   99.8%;  Pred. No. 7.4e-175;
  Matches  630;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          378 AAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCCCTA 437
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 AAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCCCTA 60

Qy          438 CAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTGAG 497
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 CAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTGAG 120

Qy          498 CCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAACTG 557
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 CCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAACTG 180

Qy          558 TTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATG 617
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 TTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATG 240

Qy          618 GAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAA 677
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 GAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAA 300

Qy          678 TCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACAC 737
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 TCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACAC 360

Qy          738 AACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTC 797
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 AACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTC 420

Qy          798 AGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATC 857
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 AGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATC 480

Qy          858 TATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATA 917
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 TATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATA 540

Qy          918 CAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAG 977
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 CAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAG 600

Qy          978 AGCAGGAGTGCTAGGCAGACTAGGTACTGGC 1008
                |||| ||||||||||||||||||||||||||
Db          601 AGCAAGAGTGCTAGGCAGACTAGGTACTGGC 631


RESULT 22
AU138097
LOCUS       AU138097                 702 bp    mRNA    linear   EST 02-AUG-2002
DEFINITION  AU138097 PLACE1 Homo sapiens cDNA clone PLACE1007839 5', mRNA
            sequence.
ACCESSION   AU138097
VERSION     AU138097.1  GI:10999618
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..702
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1007839"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="Vector: pME18SFL3"
ORIGIN

  Query Match             42.4%;  Score 628.6;  DB 1;  Length 702;
  Best Local Similarity   94.6%;  Pred. No. 1.3e-174;
  Matches  660;  Conservative    0;  Mismatches   37;  Indels    1;  Gaps    1;

Qy          633 AGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCA 692
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 AGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATAACCCA 60

Qy          693 TACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATG 752
                |||||||||||||||||||||||||||||||||||||| | ||||||||| |||||||||
Db           61 TACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAACTCCCAATG 120

Qy          753 CATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGT 812
                ||||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db          121 CATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTTGT 180

Qy          813 CTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTC 872
                |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db          181 CTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTC 240

Qy          873 ATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACC 932
                ||||||||||||||||||||||||||||||||||||| ||||||||| | ||||| || |
Db          241 ATTCTTAGTGCCCCCTATGACCATCTACACTGAACAATATTTATACAGTTATGTCATATC 300

Qy          933 TAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGG 992
                |||||||| |||||||||||||||||||||||||||||||||  ||||||||||||||||
Db          301 TAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGGAGTGCTAGG 360

Qy          993 CAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATC 1052
                   ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db          361 TGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTATC 420

Qy         1053 TCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCA 1112
                ||||||| ||||||| |||||||||| |||| | ||||||||||||||||||||||||||
Db          421 TCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTTGCAAGATCA 480

Qy         1113 ACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGC 1172
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 ACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGC 540

Qy         1173 CAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATC 1232
                  |||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||
Db          541 TGAAAGAGGGGGAACCTGTTTATTTTTAGGGGAAGAATGCTGTTATTATGTTAATCAATC 600

Qy         1233 CAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAG-AGCAGAGG 1291
                | |||| |||||||||||||||||||||||||||||| |||||||| |||  |||| |||
Db          601 CGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATNGAATACAACGTANAAGCANAGG 660

Qy         1292 AGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAA 1329
                |||||| |||||| | || |||||||||||||||||||
Db          661 AGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAA 698


RESULT 23
BX409328
LOCUS       BX409328                 924 bp    mRNA    linear   EST 01-MAY-2004
DEFINITION  BX409328 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE013YI20
            5-PRIME, mRNA sequence.
```

```
ACCESSION   BX409328
VERSION     BX409328.2  GI:46932867
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 13, 2003 this sequence version replaced gi:30652931.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAE012ZF02_AE01071_1&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1..924
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE013YI20"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             42.1%;  Score 623.8;  DB 5;  Length 924;
  Best Local Similarity   85.5%;  Pred. No. 3.8e-173;
  Matches  748;  Conservative    0;  Mismatches  120;  Indels    7;  Gaps    5;

Qy           1 ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTTCTCTTACCCCCTTTCGCTCTCACT 60
               |||| ||  || |   ||| |||||||||||||||||| |||||| ||||| ||||||||
Db          51 ATGGGCCCTCCCTTANATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT 110

Qy          61 GCACCCCCTCCATGCTGCTGTACAACCAGTAGCTCCCCTTACCAAGAGTTTCTATGAAGA 120
               ||||||||||||||| ||||||  |||||||||||||||||||||||||||||||| |||
Db         111 GCACCCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGA 170

Qy         121 ACGCGGCTTCCTGGAAATATTGATGCCCCATCATATAGGAGTTTATCTAAGGGAAACTCC 180
               | || || ||| |||||||||||||||||||| ||||||||| | |||||||||| | ||
Db         171 ATGCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 230

Qy         181 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATAACTCTGCCACTCTTTGCATG 240
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         231 ACCTTCACTGCCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 290

Qy         241 CATGCAAATACTCATTATTGGACAGGGAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 300
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db         291 CATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTT 350

Qy         301 GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAGTATGTCTGATGGGGGTGGAATT 360
               |||| ||||||||||||||||||||||||||| ||  |||||||||||||||||||| ||
Db         351 GGAGTCACTGTCTGTTGGACTTACTTCACCCAAACTGGTATGTCTGATGGGGGTGGAGTT 410

Qy         361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT 420
               |||| |||||||||||||||||| ||||| |||| |||||||||||| ||||||| ||||
Db         411 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACAT 470

Qy         421 AGCACCCCTAGCCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACC 480
```

```
                ||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         471 GGCACCTCTAGCCCCTACAAAGGACTAGATCTCTCAAAACTACATGAAACCCTCCGTACC 530

Qy         481 CATACTCGCCTGGTGAGCCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCC 540
               |||||||||||||| |||||||||||||||||||||||| |||||||||||||||| |||
Db         531 CATACTCGCCTGGTAAGCCTATTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCC 590

Qy         541 CAAAACCCTACTAACTGTTGGATGTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATC 600
               || ||||||||       ||||| ||||||||||   ||||   ||      ||   |||
Db         591 CANAACCCTACNNTACTGTGGATATGCCTCCCCCCTGACTTANNGCATATNGTT--CATC 648

Qy         601 CCTGTTCCTGAACAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTA 660
               ||||| |||| ||||||| |  ||||||||||||||||||||||||||| ||||||||||
Db         649 CCTGTACCTG-ACAATGGGAACACTTCAGCACAGAAATAAACACCACTTTCGTTTTAGTA 707

Qy         661 GGACCTCTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
               ||||||||||||||||    ||  || |||||||||| |||||| ||||| || ||||||
Db         708 GGACCTCTTGTTTCCATCTGGG--ATTACCCATACCTAAAACCTNACCTGGGT-AAATTT 764

Qy         721 AGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGA 780
               |||| |||||    | ||||||  |||||        |  ||||| | ||| |||  | |
Db         765 AGCATTACTACTTCCCCAACCACTTCCCATGCATCAGGTGGGGTACCTCCTTCCAACCAA 824

Qy         781 ATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTG 840
               |||||||||||||||||||||| |||||||| |||||| |||| | ||| |  |||||||
Db         825 ATAGTCTGCCTACCCTCAGGAA-ATTTTTTGGCTGTGGGACCTTAACCTTTGGTTGTTTG 883

Qy         841 AATGGCTCTTCAGAATCTATGTGCTTCCTCTCATT 875
               || ||||||| |||| ||||  ||   |  ||| |
Db         884 AAAGGCTCTTAAGAAACTATAGGCCCTCCTTCAAT 918


RESULT 24
BE734284
LOCUS       BE734284                 714 bp    mRNA    linear   EST 15-SEP-2000
DEFINITION  601565487F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3840572 5',
            mRNA sequence.
ACCESSION   BE734284
VERSION     BE734284.1  GI:10148276
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM530  row: m  column: 21
            High quality sequence stop: 712.
FEATURES             Location/Qualifiers
     source          1..714
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3840572"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_21"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                    of California, Berkeley) using  ZAP-cDNA synthesis kit
                    (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

  Query Match             41.2%;  Score 609.8;  DB 2;  Length 714;
  Best Local Similarity   94.3%;  Pred. No. 5e-169;
  Matches  676;  Conservative    0;  Mismatches   37;  Indels    4;  Gaps    4;

Qy        723 CAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAAT 782
              |||||||| | ||||||||| |||||||||||||||||||||||| |||||||||| |||
Db          1 CAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAAAT 60

Qy        783 AGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAA 842
              ||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db         61 AGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATC-GTGTTTGAA 119

Qy        843 TGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACAC 902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 TGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACAC 179

Qy        903 TGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCT 962
              ||||||||||||||||| | ||||| || ||||||||| ||||||||||||||||||||
Db        180 TGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCT 239

Qy        963 TCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCAC 1022
              ||| ||||||||  ||||||||||||||||   |||||||||||||| |||| |||||||
Db        240 TCC-TTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCA-TGGCGGTATCAC 297

Qy       1023 AACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGT 1082
              ||||||||||||||||||||||||||||||||||||| ||||||| |||||||||| |||
Db        298 AACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGGGT 357

Qy       1083 CACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCA 1142
              | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 CGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCA 417

Qy       1143 AAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGG 1202
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        418 AAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTAGG 477

Qy       1203 AGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAAT 1262
               |||||| ||||||||||||||||||||||| |||| ||||||||||||||||||||||||
Db        478 GGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAAT 537

Qy       1263 TCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCT 1322
              |||||||||||||||| |||||||||||||||||| |||||| | || ||||||||||||
Db        538 TCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCT 597

Qy       1323 CAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTT 1382
              ||||||||||||||||||| ||||||||||||||||||||||||||||| |||||||| |
Db        598 CAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCT 657

Qy       1383 ACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAAT 1439
              ||||||||||||||||||||||||||| ||||||||||| |||||||||||||||||
Db        658 ACTCCTCTTTGGACCCTGTATCTTTAA-CTCCTTGTTAACTTTGTCTCTTCCAGAAT 713


RESULT 25
BX430050/c
LOCUS       BX430050                 658 bp    mRNA    linear   EST 04-MAY-2004
DEFINITION  BX430050 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012YJ24
            3-PRIME, mRNA sequence.
ACCESSION   BX430050
VERSION     BX430050.2  GI:47003301
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 658)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 15, 2003 this sequence version replaced gi:30776872.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAB002ZD11_CS00181_1&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1..658
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YJ24"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             39.2%;  Score 580.4;  DB 5;  Length 658;
  Best Local Similarity   93.6%;  Pred. No. 2.6e-160;
  Matches  616;  Conservative    0;  Mismatches   41;  Indels    1;  Gaps    1;

Qy        505 AATACCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAA-ACCCTACTAACTGTTGGAT 563
              || |||| || || | |   |||| |||| |  |  |||| |||||||||||||||| ||
Db        658 AAAACCATCCACATTGGATACCATAAGGTTTTGGTACAAATACCCTACTAACTGTTGAAT 599

Qy        564 GTGCCTCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAA 623
               |||||||||||| |||||||||||||  ||||||||||||| |||||||||||||||||
Db        598 ATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAA 539

Qy        624 CTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGA 683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        538 CTTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGA 479

Qy        684 AATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAG 743
              ||||||||||||||||||||||||||||||||||||||||||||||| | |||||||||
Db        478 AATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAA 419

Qy        744 CTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAAT 803
              |||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db        418 CTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAAT 359

Qy        804 ATTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTG 863
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        358 ATTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTG 299

Qy        864 CTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATCA 923
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        298 CTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAATTA 239

Qy        924 TGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGG 983
              |||| || ||||||||| |||||||||||||||||||||||||||||||||  |||||||
Db        238 TGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGG 179

Qy        984 AGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTA 1043
              |||||||||   |||||||||||||||||| ||||||||||||||||||||||||||||
Db        178 AGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTA 119

Qy       1044 CAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTT 1103
              |||||||||||||||| ||||||| |||||||||| |||| | |||||||||||||||||
Db        118 CAAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTT 59
```

```
Qy        1104 GCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGAC 1161
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          58 GCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGAC 1

RESULT 26
BX409304
LOCUS       BX409304                 792 bp    mRNA    linear   EST 03-MAY-2004
DEFINITION  BX409304 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012YJ24
            5-PRIME, mRNA sequence.
ACCESSION   BX409304
VERSION     BX409304.2  GI:46956516
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 792)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 15, 2003 this sequence version replaced gi:30766967.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAE011ZH12_AE01057_1&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1. .792
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YJ24"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             37.5%;  Score 555.8;  DB 5;  Length 792;
  Best Local Similarity   88.0%;  Pred. No. 5.4e-153;
  Matches  610;  Conservative    0;  Mismatches   82;  Indels    1;  Gaps    1;

Qy          29 TTACTGTTCTCTTACCCCCTTTCGCTCTCACTGCACCCCCTCCATGCTGCTGTACAACCA 88
               |||||||||| |||||||  ||  ||||||||||||||||||||||| ||||||  ||||
Db           1 TTACTGTTCTTTTACCCCTCTTTACTCTCACTGCACCCCCTCCATGCCGCTGTATGACCA 60

Qy          89 GTAGCTCCCCTTACCAAGAGTTTCTATGAAGAACGCGGCTTCCTGGAAATATTGATGCCC 148
               |||||||||||||||||||||||||||| |||| || || ||| ||||||||||||||||
Db          61 GTAGCTCCCCTTACCAAGAGTTTCTATGGAGAATGCAGCGTCCCGGAAATATTGATGCCC 120

Qy         149 CATCATATAGGAGTTTATCTAAGGGAAACTCCACCTTCACTGCCCACACCCATATGCCCC 208
               |||| ||||||||| | |||||||||| | ||||||||||||||||||||||||||||||
Db         121 CATCGTATAGGAGTCTTTCTAAGGGAACCCCCACCTTCACTGCCCACACCCATATGCCCC 180

Qy         209 GCAACTGCTATAACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGGA 268
               ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |
Db         181 GCAACTGCTATCACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAA 240

Qy         269 AAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGCCACTGTCTGTTGGACTTACTTCA 328
               ||||||||||||||||||||||||||  |||||||| |||||||||||||||||||||||
Db         241 AAATGATTAATCCTAGTTGTCCTGGAACACTTGGAGTCACTGTCTGTTGGACTTACTTCA 300

Qy         329 CCCATACCAGTATGTCTGATGGGGGTGGAATTCAAGGTCAGGCAAGAGAAAAACAAGTAA 388
               |||| ||  |||||||||||||||||||| |||||| |||||||||||||||||| | ||
Db         301 CCCAAACTGGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAAGAGAAAAACATGNAA 360

Qy         389 AGGAAGCAATCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCCCTACAAAGGACTAG 448
               | |||| ||| | |||||| ||||||  |||| |||||  ||||||||||||||||||||
Db         361 AAGAAGNAATNTNCCAACTAACCCGGNTACATGGCACCTNTAGCCCCTACAAAGGACTAG 420

Qy         449 TTCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTGAGCCTATTTAATA 508
                ||||||||||||||||||||||||||| ||||||||||||||||  || ||||| ||||
Db         421 ATCTCTCAAAACTACATGAAACCCTCCGNACCCATACTCGCCTGGNAAGTCTATTNAATA 480

Qy         509 CCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAACTGTTGGATGTGCC 568
               |||||||||||   ||||||||||| ||   |||||||||||||||||| || || ||||
Db         481 CCACCCTCACTGNCCTCCATGAGGTTTCGTTCCAAAACCCTACTAACTGCTGTATATGCC 540

Qy         569 TCCCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCCTGAACAATGGAACAACTTCA 628
               |||||||| ||||||| |||||    ||||| ||||| |||||||| || |||| |||||
Db         541 TCCCCCTGAACTTCAGCCCATATGCCTCAATTCCTGTACCTGAACACTGCAACACCTTCA 600

Qy         629 GCACAGAAATAAACA-CCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATA 687
               |||||| || ||| | || |||| |||||||||||||||||||||| |||||||  ||||
Db         601 GCACAGTAACAAATACCCCCTTCTGTTTTAGTAGGACCTCTTGTTTTCAATCTGTTAATA 660

Qy         688 ACCCATACCTCAAACCTCACCTGTGTAAAATTT 720
                | | | |||||||| | ||||||||| |||||
Db         661 TCTCTTTCCTCAAACTTTACCTGTGTACAATTT 693

.
RESULT 27
BX430055/c
LOCUS       BX430055                 814 bp    mRNA    linear   EST 04-MAY-2004
DEFINITION  BX430055 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE013YI20
            3-PRIME, mRNA sequence.
ACCESSION   BX430055
VERSION     BX430055.2  GI:47006728
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 814)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 15, 2003 this sequence version replaced gi:30778885.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAB002ZG11_CS00184_1&c=4215.r
            .
FEATURES             Location/Qualifiers
     source          1. .814
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE013YI20"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
```

```
ORIGIN

  Query Match             36.3%;  Score 537.6;  DB 5;  Length 814;
  Best Local Similarity   85.3%;  Pred. No. 1.4e-147;
  Matches  671;  Conservative    0;  Mismatches  105;  Indels   11;  Gaps    6;

Qy        379 AAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACATAGCACCCCTAGCCCCTAC 438
              ||||  | ||    || ||| || ||||  | |||||  ||  ||||||    |||||
Db        814 AAACCTGGAACACTAGAAATTTCTCAACCCAACCGGGAGCAAGGCACCCTATACCCCTTA 755

Qy        439 AAAGGACTAGTTCTCTCAAAACTACATGAAACCCTCCGTACCCATACTCGCCT-GGTGAG 497
              |||||  ||   ||  |||||  |||  ||| || |||   ||  |   |||| ||| ||
Db        754 AAAGGTATAAAACTGCCAAAAAAACACTAAAACCCCCGGGGCCCAATATGCCTGGGTAAG 695

Qy        498 CCTATTTAATACCACCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCTACTAACTG 557
              || |||||| ||||||||||   ||  ||   ||| ||| ||||||||||||||||||||
Db        694 CCAATTTAAAACCACCCTCA--TGGGGCCCATAGGGCTCGGCCCAAAACCCTACTAACTG 637

Qy        558 TTGGATGTGCCTCCCCCTGCACTT---CAGGCCATACATTTCAATCCCTGTTCCTGAA-C 613
              |||||| ||||||||||   |  |     | | |    ||  ||||||||| |||||| |
Db        636 TTGGATATGCCTCCCCCCTGAACTTCAAGGCCAAATGTTTCAAATCCCTGTACCTGAACC 577

Qy        614 AATGGAACAACTT---CAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTG 670
              |||||| ||||||      || | ||  ||   | |||||||||||||||||||||||||
Db        576 AATGGANCAACTTTCAGCACAGAAAATAAACACCAACTTCCGTTTTAGTAGGACCTCTTG 517

Qy        671 TTTCCAATCTGGAAAT-AACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACT 729
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        516 TTTCCAATCTGGAAATAAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACT 457

Qy        730 ATAGACACAACCAGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGC 789
              | | ||||||||| |||||||||||||||||||||||| |||||||||| ||||||||||
Db        456 ACATACACAACCAACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGC 397

Qy        790 CTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCT 849
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        396 CTACCCTCAGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCT 337

Qy        850 TCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 TCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAA 277

Qy        910 GATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTT 969
              |||||||||||| ||||| || ||||||||| ||||||||||||||||||||||||||||
Db        276 GATTTATACAATTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTT 217

Qy        970 GTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCT 1029
              |||||  ||||||||||||||||   ||||||||||||||||||| ||||||||||||||
Db        216 GTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCT 157

Qy       1030 ACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGAC 1089
              |||||||||||||||||||||||||||||| ||||||| |||||||||| |||| | |||
Db        156 ACTCAGTTCTACTACAAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGAC 97

Qy       1090 TCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGA 1149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 TCCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGA 37

Qy       1150 AGAGCTT 1156
              |||||||
Db         36 AGAGCTT 30


RESULT 28
BE019603
LOCUS       BE019603                 586 bp    mRNA    linear   EST 06-JUN-2000
DEFINITION  ba84f03.y1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:2907101 5'
            similar to TR:095244 095244 ENVELOPE PROTEIN ;contains Alu
            repetitive element;, mRNA sequence.
ACCESSION   BE019603
VERSION     BE019603.1  GI:8279682
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
            Seq primer: -40RP from Gibco
            High quality sequence stop: 519.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2907101"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_21"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

  Query Match             36.2%;  Score 536.4;  DB 2;  Length 586;
  Best Local Similarity   94.7%;  Pred. No. 2.8e-147;
  Matches  555;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy        703 CTCACCTGTGTAAAATTTAGCAATACTATAGACACAACCAGCTCCCAATGCATCAGGTGG 762
              |||||||||||  ||||||||||||||| | ||||||||| |||||||||||||||||||
Db          1 CTCACCTGTGTTTAATTTAGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGG 60

Qy        763 GTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACC 822
              ||||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db         61 GTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTTGTCTGTGGTACC 120

Qy        823 TCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTG 882
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTG 180

Qy        883 CCCCCTATGACCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCAC 942
              ||||||||||||||||||||||||||||||||||||| | ||||| || ||||||||| |
Db        181 CCCCCTATGACCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGC 240

Qy        943 AACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGT 1002
              ||||||||||||||||||||||||||||||||  ||||||||||||||||   |||||||
Db        241 AACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGT 300

Qy       1003 ACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAAATA 1062
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||| ||
Db        301 ACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAACTA 360

Qy       1063 AATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCC 1122
              ||||| |||||||||| |||| | ||||||||||||||||||||||||||||||||||||
Db        361 AATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTCC 420

Qy       1123 CTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGG 1182
              ||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db        421 CTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGG 480
```

```
Qy       1183 GGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTC 1242
              |||||||||||||||||||| |||||| ||||||||||||||||||||||| |||| |||
Db        481 GGAACCTGTTTATTTTTAGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTC 540

Qy       1243 ACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAG 1288
              |||||||||||||||||||||||||||||||||||| |||||||||
Db        541 ACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAACGTAGAGCAG 586


RESULT 29
BG572445
LOCUS       BG572445                 790 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602593490F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4720846 5',
            mRNA sequence.
ACCESSION   BG572445
VERSION     BG572445.1  GI:13580098
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 790)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1576  row: c  column: 23
            High quality sequence stop: 624.
FEATURES             Location/Qualifiers
     source          1. .790
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4720846"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_79"
                     /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI
                     (ggccattatggcc); 5' and 3' adaptors were used in cloning
                     as follows: 5' adaptor sequence:  5'-CACGGCCATTATGGCC-3'
                     and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.3
                     kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             36.1%;  Score 535.2;  DB 4;  Length 790;
  Best Local Similarity   90.5%;  Pred. No. 7e-147;
  Matches  650;  Conservative    0;  Mismatches   58;  Indels   10;  Gaps    7;

Qy        742 AGCTCCCAATGCATCAGGTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTCAGGA 801
              | |||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||
Db          1 AACTCCCAATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGA 60

Qy        802 ATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTATG 861
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 ATATTTTTTGTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATG 120

Qy        862 TGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAAT 921
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        121 TGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAGT 180
```

```
Qy        922 CATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCAGAGCA 981
               ||||| || ||||||||| |||||||||||||||||||||||||||||||||  |||||
Db        181 TATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTTATAGGAGCA 240

Qy        982 GGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACCTCTACTCAGTTCTAC 1041
              |||||||||||   ||||||||||||||||||| ||||||||||||||||||||||||||
Db        241 GGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTAC 300

Qy       1042 TACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACC 1101
              |||||||||||||||||| ||||||| |||||||||| |||| | |||||||||||||||
Db        301 TACAAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACC 360

Qy       1102 TTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCA-AAATCGAAGAGCTTTAGA 1160
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        361 TTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCACAAATCGAAGAGCTTTAGA 420

Qy       1161 CTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGTTATTA 1220
              ||||||||||||  |||||||||||||||||||||||||||| |||||| ||||||||||
Db        421 CTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTTAGGGGAAGAATGCTGTTATTA 480

Qy       1221 TGTTAATCAATCCAGAATTGTCACTGAGAAAGTTA-AAGAAATTCGAGATCG--AATACA 1277
              ||||||||||||| |||| |||||||||||||||| ||||||||||||||||  |
Db        481 TGTTAATCAATCCGGAATCGTCACTGAGAAAGTTACAAGAAATTCGAGATCGCAATACAC 540

Qy       1278 ATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATGCC 1337
              | |||||||||||||||||| |||||| | || |||||||   |     |||||||||||
Db        541 ACGTAGAGCAGAGGAGCTTCGAAACACTGCACCCTGGGGC--TCCTCAGCAATGGATGCC 598

Qy       1338 CTGGGTTCTCCCCTTCTTAGGACCTCTAGC-AGCTCTAATATTGTTACTCCTCTTTGGAC 1396
                 || |   |||||||||||||||||||| |||| || ||||| |||||||| |||||
Db        599 --TGGATTCTCCCTTCTTAGGACCTCTAGCAAGCTATACTATTGCTACTCCTC-TTGGAG 655

Qy       1397 CCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCT 1454
              |||||||||||||||||||||| || |  |||||||||||||| |||| |||||| ||
Db        656 CCTGTATCTTTAACCTCCTTGTCAACTGAGTCTCTTCCAGAATCGAAGGTGTAAACCT 713


RESULT 30
BX450495/c
LOCUS       BX450495                 760 bp    mRNA    linear   EST 05-MAY-2004
DEFINITION  BX450495 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE013YI20
            3-PRIME, mRNA sequence.
ACCESSION   BX450495
VERSION     BX450495.2  GI:47038997
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 22, 2003 this sequence version replaced gi:31018916.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AU006ZG02_U0495_1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE013YI20"
                     /tissue_type="PLACENTA"
```

```
                    /clone_lib="Homo sapiens PLACENTA"
                    /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                    with a NotI-oligo(dT) primer. Five prime end enriched,
                    double-strand cDNA was digested with Not I and cloned into
                    the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                    Library was not normalized."
ORIGIN

  Query Match             32.3%;  Score 478.6;  DB 5;  Length 760;
  Best Local Similarity   83.3%;  Pred. No. 4.3e-130;
  Matches  604;  Conservative    0;  Mismatches  114;  Indels    7;  Gaps    5;

Qy         682 GAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACT-ATAGACACAAC 740
               | || ||||||  || |   ||||  | | | ||||| | || || ||  |  |  ||||
Db         727 GCAACAACCCAAGCCCCCTCCCTCGGCGGGGCAAAATCTCGCTATTCTCCTCTAATCAAC 668

Qy         741 CAGCTCCCAATGCATCAG---GTGGGTAACACCTCCCACACGAATAGTCTGCCTACCCTC 797
               |  |  || ||| |||||    | |     |  ||| ||||  |    |     ||||  |
Db         667 CGACCACCCATGGATCAGGGTGAGTCGTCCCTCTCTCACAAAATAGTTTGCTATACCTCC 608

Qy         798 AGGAATATTTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATC 857
               || |||| |  |||| || ||  |||||||| | | | ||  |||| |||||||||||| |
Db         607 AGAAATAGTGGTTGTGTGAGGCTCCTCAGCCAACCGTAGTATGAAGGGCTCTTCAGAAAC 548

Qy         858 TATGTGCTTCCTCTCATTCTTAGTGCCCCCTATGACCATCTACACTGAAC-AAGATTTAT 916
                |||||| ||||| || ||| || |||||||||||||||||| ||||||||| |||||||| 
Db         547 AATGTGC-TCCTCGCAATCTAAGAGCCCCCTATGACCATCGACACTGAACTAAGATTTAG 489

Qy         917 ACAATCATGTCGTACCTAAGCCCCACAACAAAAGAGTACCCATTCTTCCTTTTGTTATCA 976
                | || |||||| || ||||||||  ||||||||| ||  ||||||||||||||||||||  
Db         488 CCCATTATGTCATATCTAAGCCCAGCAACAAAAAAGCTCCCATTCTTCCTTTTGTTATAG 429

Qy         977 GAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAA-CCTCTACTCAG 1035
                |||||||||||||||   |||||| |||||||||| | |||||||| |||||||||||
Db         428 AAGCAGGAGTGCTAGGAGCACTAGGAACTGGCATTGTCGGTATCACAAGCCTCTACTCAG 369

Qy        1036 TTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTG 1095
               |||||||||  ||||||||||||| ||||||| ||||||||||| |||| |  |||||||||
Db         368 TTCTACTACGGACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTG 309

Qy        1096 GTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCT 1155
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         308 GTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCT 249

Qy        1156 TTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTAGGAGAAGAACGCTGT 1215
               |||||||||||||||||  |||||||||| ||||||||||||||||| |||||| |||||
Db         248 TTAGACTTGCTAACCGCTGAAAGAGGGGGGACCTGTTTATTTTTAGGGGAAGAATGCTGT 189

Qy        1216 TATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATA 1275
               || ||||||||||||||| |||| ||||||||||||||||||||||||||||||||||||
Db         188 TACTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATA 129

Qy        1276 CAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCTCCTCAGCCAATGGATG 1335
               ||| |||||||||||||||||| |||||| | || |||||||||||||||||||||||||
Db         128 CAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATG 69

Qy        1336 CCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCCTCTTTGGA 1395
               |||||| ||||||||||||||||||||||||||||||| ||||||||| ||||||||||||||
Db          68 CCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTGGA 9

Qy        1396 CCCTG 1400
               |||||
Db           8 CCCTG 4

Search completed: January 22, 2005, 19:31:41
Job time : 4420.62 secs
```

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2005  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         January 22, 2005, 20:15:26 ; Search time 114 Seconds
                                          (without alignments)
                                          1551.347 Million cell updates/sec

Title:          US-09-319-156B-10
Perfect score:  2634
Sequence:       1 MALPYHTFLFTVLLPPFALT..........KFVSSRIEAVKLQMVLQMEP 493

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters:      2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_23Sep04:*
                1:  geneseqp1980s:*
                2:  geneseqp1990s:*
                3:  geneseqp2000s:*
                4:  geneseqp2001s:*
                5:  geneseqp2002s:*
                6:  geneseqp2003as:*
                7:  geneseqp2003bs:*
                8:  geneseqp2004s:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                  %
Result           Query
  No.   Score    Match Length DB  ID                     Description
-----------------------------------------------------------------------------
    1    2632    99.9    493  2  AAW71068               Aaw71068 Multiple
    2    2632    99.9    493  2  AAW99553               Aaw99553 Protein e
    3    2632    99.9    493  8  ADG14847               Adg14847 MSRV asso
    4  2622.5    99.6    541  3  AAB08195               Aab08195 Amino aci
    5    2613    99.2    542  3  AAB19069               Aab19069 Amino aci
    6    2605    98.9    493  7  ADB84401               Adb84401 MSRV-1 as
    7    2447    92.9    542  5  AAE25054               Aae25054 Human EMB
    8    2319    88.0    538  2  AAY08622               Aay08622 Human sec
    9    2319    88.0    538  3  AAY67313               Aay67313 Human sec
   10    2319    88.0    538  5  AAE14540               Aae14540 Human syn
   11    2319    88.0    538  7  ADC38777               Adc38777 Human sec
   12    2314    87.9    538  4  AAB75138               Aab75138 HERV-W en
   13    2314    87.9    538  4  AAB67652               Aab67652 Amino aci
   14    2314    87.9    540  3  ABP40980               Abp40980 Human ret
   15    2314    87.9    685  3  ABP40988               Abp40988 Human ret
   16    2314    87.9    846  3  ABP41050               Abp41050 HERV-7q e
   17  2308.5    87.6    544  7  ADE09243               Ade09243 Novel pro
   18  2308.5    87.6    544  7  ADF60170               Adf60170 Human con
   19    2237    84.9    538  2  AAW99345               Aaw99345 Human end
   20  1235.5    46.9    263  4  AAM18008               Aam18008 Peptide #
   21  1235.5    46.9    263  4  ABB37043               Abb37043 Peptide #
   22  1235.5    46.9    263  4  AAM30519               Aam30519 Peptide #
   23  1235.5    46.9    263  4  ABB31812               Abb31812 Peptide #
   24  1235.5    46.9    263  4  ABB22358               Abb22358 Protein #
   25  1235.5    46.9    263  4  AAM70185               Aam70185 Human bon
```

```
   26  1235.5    46.9    263  4  AAM57770               Aam57770 Human bra
   27  1235.5    46.9    263  4  ABG51885               Abg51885 Human liv
   28  1235.5    46.9    263  4  AAM05648               Aam05648 Peptide #
   29  1235.5    46.9    263  5  ABG39819               Abg39819 Human pep
   30  1130.5    42.9    446  4  ABG07532               Abg07532 Novel hum
   31  1130.5    42.9    446  4  ABG20007               Abg20007 Novel hum
   32    1122    42.6    531  5  ABP63036               Abp63036 Human pol
   33   965.5    36.7    368  4  ABG28311               Abg28311 Novel hum
   34   965.5    36.7    633  4  ABG28306               Abg28306 Novel hum
   35     591    22.4    120  4  AAB75137               Aab75137 HERV-W tr
   36     589    22.4    122  5  ADK35467               Adk35467 Novel hum
   37   571.5    21.7    563  7  ADG86762               Adg86762 Chimeric
   38   551.5    20.9    567  2  AAR71700               Aar71700 Spleen ne
   39     517    19.6    162  7  ADB84404               Adb84404 MSRV-1 as
   40     516    19.6    162  2  AAW71069               Aaw71069 Multiple
   41     516    19.6    162  2  AAW99554               Aaw99554 Protein e
   42     516    19.6    162  8  ADG14850               Adg14850 MSRV asso
   43   514.5    19.5    584  4  ABG22871               Abg22871 Novel hum
   44   514.5    19.5    584  4  ABG19851               Abg19851 Novel hum
   45   514.5    19.5    584  4  ABG05606               Abg05606 Novel hum

                              ALIGNMENTS

RESULT 1
AAW71068
ID   AAW71068 standard; protein; 493 AA.
XX
AC   AAW71068;
XX
DT   29-DEC-1998  (first entry)
XX
DE   Multiple sclerosis associated retrovirus protein 3.
XX
KW   Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW   env gene; rheumatoid arthritis-associated virus.
XX
OS   Multiple sclerosis associated retrovirus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 39
FT                   /note= "Encoded by tga"
XX
PN   WO9823755-A1.
XX
PD   04-JUN-1998.
XX
PF   26-NOV-1997;   97WO-IB001482.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B;
XX
DR   WPI; 1998-322732/28.
DR   N-PSDB; AAV43217.
XX
PT   New nucleic acid from retroviruses - useful for diagnosis, prevention and
PT   treatment of, e.g. multiple sclerosis.
XX
PS   Disclosure; Page 185-187; 286pp; English.
XX
CC   The present sequence represents a multiple sclerosis (MS) associated
CC   retrovirus (MSRV) protein used in the method of the invention. The
CC   invention provides complete or partial genomic sequences of the MSRV-1
CC   pol gene, gag gene and env gene, and polypeptides encoded by these genes.
CC   The invention also provides antibodies raised against the polypeptides.
CC   The genomic sequences, polypeptides and antibodies are also claimed
CC   useful for diagnosing infection by MS and rheumatoid arthritis-associated
CC   viruses, and also for prevention and treatment of infection with these
CC   viruses
XX
SQ   Sequence 493 AA;

  Query Match             99.9%;  Score 2632;  DB 2;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-229;
  Matches  493;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              |||||||||||||
Db        481 EAVKLQMVLQMEP 493


RESULT 2
AAW99553
ID   AAW99553 standard; protein; 493 AA.
XX
AC   AAW99553;
XX
DT   17-OCT-2003  (revised)
DT   27-AUG-2003  (revised)
DT   08-JUN-1999  (first entry)
XX
DE   Protein encoded by clone C15 from MSRV-1.
XX
KW   Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KW   rheumatoid polyarthritis.
XX
OS   Viruses.
XX
FH   Key               Location/Qualifiers
FT   Misc-difference   39
FT                     /label= unknown
XX
PN   FR2765588-A1.
XX
PD   08-JAN-1999.
XX
PF   07-JUL-1997;   97FR-00008816.
XX
PR   07-JUL-1997;   97FR-00008816.
XX
PA   (INMR ) BIO MERIEUX.
XX
DR   WPI; 1999-098275/09.
DR   N-PSDB; AAX29703.
XX
PT   Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT   multiple sclerosis or rheumatoid polyarthritis.
XX
PS   Claim 2; Page 38-39; 83pp; French.
XX
CC   This sequence represents the protein encoded by clone C15 from a novel
CC   multiple sclerosis related virus type 1 (MSRV1). The sequence can be used
CC   in diagnostic, prophylactic or therapeutic compositions to inhibit
CC   expression of a multiple sclerosis related virus and/or virus associated
CC   with rheumatoid polyarthritis. (Updated on 27-AUG-2003 to correct OS
CC   field.) (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 493 AA;

  Query Match             99.9%;  Score 2632;  DB 2;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-229;
  Matches  493;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              |||||||||||||
Db        481 EAVKLQMVLQMEP 493


RESULT 3
ADG14847
ID   ADG14847 standard; protein; 493 AA.
XX
AC   ADG14847;
XX
DT   26-FEB-2004  (first entry)
XX
DE   MSRV associated protein #7.
XX
KW   pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX
```

```
OS   Multiple sclerosis associated retrovirus.
XX
PN   US2003198647-A1.
XX
PD   23-OCT-2003.
XX
PF   03-APR-2002; 2002US-00114104.
XX
PR   26-NOV-1996;   96US-00756429.
PR   26-NOV-1997;   97US-00979847.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2004-032461/03.
XX
PT   New isolated nucleic acid and their fragments having the pol gene of a
PT   retrovirus, useful for diagnosing, preventing and/or treating multiple
PT   sclerosis and/or rheumatoid arthritis.
XX
PS   Disclosure; SEQ ID NO 106; 193pp; English.
XX
CC   The invention relates to an isolated nucleic acid which comprises the pol
CC   gene of a retrovirus associated with multiple sclerosis or rheumatoid
CC   arthritis. The methods and compositions of the present invention are
CC   useful for diagnosing, preventing and/or treating multiple sclerosis
CC   and/or rheumatoid arthritis. The present sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence 493 AA;

  Query Match             99.9%;  Score 2632;  DB 8;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-229;
  Matches  493;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              |||||||||||||
Db        481 EAVKLQMVLQMEP 493
```

```
RESULT 4
AAB08195
ID   AAB08195 standard; protein; 541 AA.
XX
AC   AAB08195;
XX
DT   06-AUG-2003  (revised)
DT   04-DEC-2000  (first entry)
XX
DE   Amino acid sequence of the MSRV-1 3' env and LTR regions.
XX
KW   MSRV-1; pol region; long terminal repeat; LTR; RU5 region; retrovirus.
XX
OS   Multiple sclerosis associated retrovirus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .27
FT                   /note= "signal peptide"
FT   Modified-site   58. .60
FT                   /note= "N-linked glycosylation site"
FT   Modified-site   92. .94
FT                   /note= "N-linked glycosylation site"
FT   Modified-site   213. .215
FT                   /note= "N-linked glycosylation site"
FT   Modified-site   233. .235
FT                   /note= "N-linked glycosylation site"
FT   Modified-site   280. .282
FT                   /note= "N-linked glycosylation site"
FT   Cleavage-site   289. .290
FT                   /note= "putative cleavage site"
FT   Cleavage-site   315. .316
FT                   /note= "putative cleavage site"
FT   Modified-site   408. .410
FT                   /note= "N-linked glycosylation site"
XX
PN   WO200047745-A1.
XX
PD   17-AUG-2000.
XX
PF   15-FEB-2000; 2000WO-IB000159.
XX
PR   15-FEB-1999;   99EP-00420041.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Perron H,  Komurian-Pradel F;
XX
DR   WPI; 2000-506097/45.
DR   N-PSDB; AAA63826.
XX
PT   Nucleotide fragment of LTR-RU5 region from Multiple Sclerosis retrovirus
PT   (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological
PT   sample.
XX
PS   Disclosure; Fig 2; 23pp; English.
XX
CC   The present sequence is encoded by a nucleotide sequence corresponding
CC   to the 3' env region and long terminal repeat sequences from clone CL6 of
CC   Multiple ~~Sclerosis~~ retrovirus (MSRV-1). The specification describes a
CC   long terminal repeat (LTR)-RU5 region which encodes the expression of a
CC   MSRV-1 protein. This is unusual for LTRs, in particular in the RU5
CC   region. Probes and antibodies to the MSRV-1 retrovirus protein and
CC   encoding polynucleotide sequences are used to detect the presence of MSRV
CC   -1 retrovirus in a biological sample. (Updated on 06-AUG-2003 to correct
CC   OS field.)
XX
SQ   Sequence 541 AA;

  Query Match             99.6%;  Score 2622.5;  DB 3;  Length 541;
  Best Local Similarity   99.8%;  Pred. No. 3.7e-228;
  Matches  492;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;
```

```
Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFL-RTRLPGNIDAPSYRSLSKGNS 59

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 119

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 179

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 239

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 299

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 359

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 419

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 479

Qy        481 EAVKLQMVLQMEP 493
              |||||||||||||
Db        480 EAVKLQMVLQMEP 492
```

```
RESULT 5
AAB19069
ID   AAB19069 standard; protein; 542 AA.
XX
AC   AAB19069;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Amino acid sequence of an envelope (env) protein of MSRV-1.
XX
KW   Envelope protein; MSRV-1; superantigen; autoimmune disease; Vbeta16;
KW   Vbeta17; multiple sclerosis; vaccine.
XX
OS   Human endogenous retrovirus.
XX
PN   WO200057185-A1.
XX
PD   28-SEP-2000.
XX
PF   20-MAR-2000; 2000WO-FR000691.
XX
PR   19-MAR-1999;    99FR-00003622.
PR   28-OCT-1999;    99FR-00013755.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Lafont M;
XX
DR   WPI; 2000-638279/61.
DR   N-PSDB; AAA96625.
XX
PT   Detecting superantigen activity, useful for identifying agents for
PT   treatment or prevention of autoimmune disease, from expansion or loss of
PT   particular lymphocyte Vbeta determinants.
XX
PS   Claim 28; Page 124-126; 134pp; French.
XX
CC   The present sequence represents an envelope protein of MSRV-1. The
CC   envelope protein expressed by the endogenous human retrovirus MRSV-1 has
CC   superantigen activity associated with autoimmune disease. The protien can
CC   be detected using the method of the invention. The specification
CC   describes a process for detecting activity of a superantigen in a
CC   biological sample. The process comprises identifying large scale
CC   expansion or loss of lymphocytes that carry at least one of the Vbeta16
CC   and/or Vbeta17 determinants. The method is used to screen for agents that
CC   inhibit the superantigen, especially those associated with MRSV-1 which
CC   is implicated in autoimmune disease, particularly multiple sclerosis.
CC   These agents are potentially useful for treatment or prevention (e.g. as
CC   vaccines). of autoimmune diseases
XX
SQ   Sequence 542 AA;

  Query Match             99.2%;  Score 2613;  DB 3;  Length 542;
  Best Local Similarity   98.8%;  Pred. No. 2.7e-227;
  Matches  487;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLWRTRLPGNIDAPSYRSLSKGNS 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||| |||||||||:| |||||||||| |||||||||
Db        421 IRDRIQCRAEELQNTERWGLLSQWMPWTLPFLGPLAAIIFLLLFGPCIFNFLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              ||||||:||||||
Db        481 EAVKLQIVLQMEP 493

RESULT 6
ADB84401
ID   ADB84401 standard; protein; 493 AA.
XX
AC   ADB84401;
XX
DT   04-DEC-2003  (first entry)
XX
DE   MSRV-1 associated protein/peptide sequence #10.
XX
KW   multiple sclerosis; rheumatoid arthritis; gag; pol;
KW   reverse transcriptase; ribonuclease H.
XX
OS   Unidentified;.
XX
PN   US2003039664-A1.
```

```
XX
PD   27-FEB-2003.
XX
PF   26-NOV-1997;   97US-00979847.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (PERR/) PERRON H.
PA   (BESE/) BESEME F.
PA   (BEDI/) BEDIN F.
PA   (PARA/) PARANHOS-BACCALA G.
PA   (KOMU/) KOMURIAN-PRADEL F.
PA   (JOLI/) JOLIVET-REYNAUD C.
PA   (MAND/) MANDRAND B.
PA   (GARS/) GARSON J A.
PA   (TUKE/) TUKE P W.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2003-512253/48.
XX
PT   New isolated or purified nucleic acid associated with multiple sclerosis
PT   and/or rheumatoid arthritis, useful for detecting a virus associated with
PT   multiple sclerosis or rheumatoid arthritis in a biological sample.
XX
PS   Claim 31; Page 78-80; 193pp; English.
XX
CC   The invention relates to an isolated or purified nucleic acid from a
CC   virus associated with multiple sclerosis and/or rheumatoid arthritis,
CC   multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
CC   pol, gag or reverse transcriptase genes (or their fragments) encoding the
CC   proteins or defined peptides (including immunodominant peptides,
CC   antigenic peptides or conserved motifs).Also included are a process for
CC   detecting a virus associated with multiple sclerosis or rheumatoid
CC   arthritis in a biological sample, a nucleic acid probe for the detection
CC   of a virus associated with multiple sclerosis or rheumatoid arthritis, a
CC   primer for the amplification by polymerisation of a nucleic acid of a
CC   viral material associated with multiple sclerosis or rheumatoid
CC   arthritis, a polypeptide exhibiting an inhibitory activity on the
CC   proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC   and an antibody directed against the MSRV-1 virus obtained by
CC   immunologically reacting a human or animal body or cells with an
CC   immunogenic agent consisting of the antigenic polypeptide defined above.
CC   The nucleic acids are useful for detecting a biological sample a virus
CC   associated with multiple sclerosis or rheumatoid arthritis, or for
CC   detecting in a biological sample, the presence of or exposure to a virus
CC   associated with multiple sclerosis or rheumatoid arthritis. The present
CC   sequence is a claimed MSRV-associated sequence whose identity cannot be
CC   accurately determined. Note: The SEQ ID numbers for the sequences as
CC   displayed in the main body of the patent do not match the SEQ ID numbers
CC   in the sequence listing. Consequently those sequences mentioned in the
CC   claims may not be the sequences the authors intended to claim.
XX
SQ   Sequence 493 AA;

  Query Match             98.9%;  Score 2605;  DB 7;  Length 493;
  Best Local Similarity   98.8%;  Pred. No. 1.3e-226;
  Matches  487;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              |||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGQAREKZVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLHFRPYISIPVPEQWNXFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||| |||||||||||||||||||||||||||||||| ||||||||||||:||||||||||
Db        301 TEQXLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTZIGSITTSTQFYYKLSZEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTZKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              |||||||||||||
Db        481 EAVKLQMVLQMEP 493


RESULT 7
AAE25054
ID   AAE25054 standard; protein; 542 AA.
XX
AC   AAE25054;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Human EMBRY-2 protein.
XX
KW   Human; embryogenesis associated protein; AIDS; reproductive disorder;
KW   infertility; endometriosis; endometrial tumour; inflammatory disorder;
KW   autoimmune disorder; acquired immune deficiency syndrome; transgenic;
KW   ovarian tumour; contact dermatitis; placenta disorder; preeclampsia;
KW   EMBRY-2; allergy; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .20
FT                   /label= Signal_peptide
FT   Protein         21. .542
FT                   /note= "Mature EMBRY-2 protein"
FT   Domain          263. .484
FT                   /note= "ENV polyprotein (coat polyprotein) domain"
FT   Domain          455. .473
FT                   /note= "Transmembrane domain"
XX
PN   WO200248362-A2.
XX
PD   20-JUN-2002.
XX
PF   14-NOV-2001; 2001WO-US043956.
XX
PR   15-NOV-2000; 2000US-0249407P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Ramkumar J,  Arvizu C;
XX
DR   WPI; 2002-537629/57.
DR   N-PSDB; AAD41225.
XX
PT   New polypeptides of human embryogenesis associated proteins for screening
PT   modulators useful for treating or preventing disorders e.g.
PT   endometriosis, infertility, allergy, preeclampsia.
XX
PS   Claim 57; Page 94-95; 97pp; English.
XX
```

```
CC   The invention relates to human embryogenesis associated proteins (EMBRY)
CC   and nucleic acid molecules encoding such proteins. EMBRY sequences are
CC   useful for screening modulators useful for treating or preventing
CC   disorders associated with abnormal expression of EMBRY. The disorders
CC   treated include reproductive disorders such as infertility,
CC   endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory
CC   disorder such as acquired immune deficiency syndrome (AIDS), allergies,
CC   contact dermatitis; disorders of the placenta such as preeclampsia,
CC   abruptio placentae etc. Sequences of the invention are also useful for
CC   analysing a proteome of a tissue or a cell type. EMBRY proteins are
CC   useful as immunogens for preparing antibodies. Polynucleotides of the
CC   invention are useful for creating knockin humanised animals or transgenic
CC   animals to model human diseases. They are also used in gene therapy. The
CC   present sequence is human EMBRY-2 protein
XX
SQ   Sequence 542 AA;

  Query Match             92.9%;  Score 2447;  DB 5;  Length 542;
  Best Local Similarity   93.1%;  Pred. No. 2.9e-212;
  Matches  459;  Conservative    8;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||  |||||| |||:|||| || | ||||||| || | | | :|||||||||||||
Db          1 MALPYCIFLFTVLSPPFSLTAPSPCHCRTSSSPYQAFLWRMRRPRHIDAPSYRSLSKGNP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
               |||||||| |||||||||||||||||||||||||||||||||:|||||||| ||||||:
Db         61 AFTAHTHMPHNCYNSATLCMHANTHYWTGKMINPSCPGGLGATICWTYFTHTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| |||||||||| |||||||||||| |||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHSTPSPYKGLDLSKLHETLRTHTHLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||||||||||||||:|||||||||||||||||:||||||||||| ||||||||||
Db        181 QNPTNCWMCLPLHFRPYISVPVPEQWNNFSTEINTTSILVGPLVSNLEIIHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||| ||:|||||||||||:||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIATTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEECCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        421 IRDRIQCRAEELQNTEHWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              |||||||||||||
Db        481 EAVKLQMVLQMEP 493


RESULT 8
AAY08622
ID   AAY08622 standard; protein; 538 AA.
XX
AC   AAY08622;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human secreted protein AJ172_2.
XX
KW   Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW   bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW   cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW   haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;
KW   haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW   cadherin; tumour invasion suppressor; gene therapy; tissue growth.
XX
OS   Homo sapiens.
XX
PN   WO9926972-A1.
XX
PD   03-JUN-1999.
XX
PF   17-NOV-1998;   98WO-US024614.
XX
PR   21-NOV-1997;   97US-00976110.
PR   18-MAY-1998;   98US-00080478.
PR   20-OCT-1998;   98US-00175928.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M;
XX
DR   WPI; 1999-357813/30.
DR   N-PSDB; AAX77526.
XX
PT   New polynucleotides encoding secreted proteins.
XX
PS   Claim 14a; Page 101-103; 142pp; English.
XX
CC   This invention describes novel human secreted proteins encoded by
CC   polynucleotides isolated from human adult testes, adult brain, adult
CC   blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC   libraries. The products of the invention are predicted to have biological
CC   activities which would make them suitable for treating, preventing or
CC   ameliorating medical conditions in humans and animals, although no
CC   supporting data is given. Suggested activities include nutritional
CC   activity, cytokine and cell proliferation/differentiation activity,
CC   immune stimulating (e.g. as vaccines) or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, and tumour
CC   inhibition activity. The polynucleotides are also stated to be useful for
CC   gene therapy
XX
SQ   Sequence 538 AA;

  Query Match             88.0%;  Score 2319;  DB 2;  Length 538;
  Best Local Similarity   88.6%;  Pred. No. 1.1e-200;
  Matches  436;  Conservative   18;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
```

```
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492


RESULT 9
AAY67313
ID   AAY67313 standard; protein; 538 AA.
XX
AC   AAY67313;
XX
DT   11-APR-2000  (first entry)
XX
DE   Human secreted protein AJ172_2 amino acid sequence.
XX
KW   Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW   placental pathology; metastasis inhibition; nutritional activity;
KW   immune stimulator; haematopoiesis regulator; tissue growth;
KW   tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115.
XX
OS   Homo sapiens.
XX
PN   WO9960020-A1.
XX
PD   25-NOV-1999.
XX
PF   17-MAY-1999;   99WO-US010915.
XX
PR   18-MAY-1998;   98US-00080478.
PR   20-OCT-1998;   98US-00175928.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Mi S,  Treacy M;
XX
DR   WPI; 2000-116311/10.
DR   N-PSDB; AAZ59468.
XX
PT   New polynucleotides encoding secreted cDNA libraries, used to develop
PT   products for the diagnosis and treatment of neoplastic disease.
XX
PS   Claim 15; Page 108-110; 149pp; English.
XX
CC   This is the human secreted protein AJ172_2 amino acid sequence, the
CC   polynucleotide encoding AJ172_2 was obtained from a human adult testes
CC   cDNA library. The invention relates to secreted human and murine
CC   proteins. The polynucleotides and proteins are predicted to have
CC   biological activities which would make them suitable for treating,
CC   preventing or ameliorating medical conditions in humans and animals.
CC   Detection of the levels of the proteins can be used for the diagnosis of
CC   e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate
CC   the expression or function of the proteins may be used for treating a
CC   neoplastic disease and inhibiting metastasis. Other suggested activities
CC   include nutritional activity (e.g. in feeds), cytokine and cell
CC   proliferation/differentiation activity, immune stimulating (e.g. as
CC   vaccines) or suppressing activity, haematopoiesis regulating activity,
CC   tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC   invasion suppressor activity, and tumour inhibition activity. The
CC   polynucleotide sequences are also stated to be useful for gene therapy
XX
SQ   Sequence 538 AA;

  Query Match             88.0%;  Score 2319;  DB 3;  Length 538;
  Best Local Similarity   88.6%;  Pred. No. 1.1e-200;
  Matches  436;  Conservative   18;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492


RESULT 10
AAE14540
ID   AAE14540 standard; protein; 538 AA.
XX
AC   AAE14540;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Human syncytin protein.
XX
KW   Human; syncytin; preeclampsia; gestational trophoblast disorder;
KW   choriocarcinoma; hydatiform mole; placental site tumour; abortion;
KW   envelope gene; human endogenous defective retrovirus; HERV-W.
XX
OS   Homo sapiens.
XX
PN   WO200204678-A2.
XX
PD   17-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-US021719.
XX
PR   07-JUL-2000; 2000US-0216657P.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Keith JC,  Mccoy JM,  Mi S;
XX
DR   WPI; 2002-171727/22.
DR   N-PSDB; AAD24195.
XX
PT   Identifying a compound for treating a subject with or at risk of
PT   developing preeclampsia, comprises determining whether the expression or
PT   activity of syncytin in the cell is modulated in the presence of a test
PT   compound.
XX
PS   Disclosure; Page 42-43; 43pp; English.
XX
CC   The invention relates to identifying compounds which are modulators of
CC   syncytin expression. The syncytin modulators are useful in diagnosis and
CC   treatment of preeclampsia and gestational trophoblast disorders (e.g.
CC   choriocarcinoma, hydatiform mole, placental site tumour and missed/
CC   incomplete abortion). Syncytin is a human gene derived from the envelope
CC   gene of human endogenous defective retrovirus, HERV-W. The present
CC   invention is based partly on the discovery that syncytin expression is
CC   dramatically reduced in preeclampsia, and is also mis-localised to the
CC   apical syncytiotrophoblast membrane. The present sequence is human
CC   syncytin protein
XX
SQ   Sequence 538 AA;

  Query Match             88.0%;  Score 2319;  DB 5;  Length 538;
  Best Local Similarity   88.6%;  Pred. No. 1.1e-200;
  Matches  436;  Conservative   18;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492


RESULT 11
ADC38777
ID   ADC38777 standard; protein; 538 AA.
XX
AC   ADC38777;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human secreted protein #63.
XX
KW   immune disorder; severe combined immunodeficiency; SCID;
KW   autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW   rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;
KW   lymphoid cell deficiency; osteoporosis; osteoarthritis;
KW   peripheral nervous system disease; peripheral neuropathy;
KW   Alzheimer's disease; Parkinson's disease; coagulation disorder;
KW   inflammatory disease; systemic inflammatory response syndrome; SIRS;
KW   ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;
KW   hypersensitivity; regeneration; neural cell proliferation; fertility;
KW   tumour; chemokine; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   US2002193567-A1.
XX
PD   19-DEC-2002.
XX
PF   02-APR-2002; 2002US-00114893.
XX
PR   11-AUG-1995;   95US-00514014.
PR   05-APR-1996;   96US-00628364.
PR   19-APR-1996;   96US-00635311.
PR   07-JUN-1996;   96US-00659224.
PR   17-JUN-1996;   96US-00664596.
PR   09-JUL-1996;   96US-00677231.
PR   26-JUL-1996;   96US-00686878.
PR   23-AUG-1996;   96US-00701819.
PR   27-SEP-1996;   96US-00721488.
PR   27-SEP-1996;   96US-00721798.
PR   27-SEP-1996;   96US-00721923.
PR   27-SEP-1996;   96US-00721926.
PR   25-OCT-1996;   96US-00738367.
PR   30-OCT-1996;   96US-00739775.
PR   13-JAN-1997;   97US-00783395.
PR   10-APR-1997;   97US-00833823.
PR   02-JUN-1997;   97US-00867677.
PR   05-SEP-1997;   97US-00924838.
PR   06-OCT-1999;   99US-00413232.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR,  Spaulding V,  Carlin-Duckett M;
PI   Kelleher K;
XX
DR   WPI; 2003-657236/62.
DR   N-PSDB; ADC38776.
XX
PT   Proteins AZ3021 encoded by clone AZ3021 from human adult colon, and
PT   BD12716 encoded by clone BD12716 from human fetal kidney cDNA library,
PT   useful for treating e.g. multiple sclerosis and rheumatoid arthritis.
XX
PS   Disclosure; SEQ ID NO 135; 412pp; English.
XX
CC   The invention relates to a protein comprising fully defined AZ302 1
CC   protein or BD127 1 6 protein. The polynucleotides are useful for
CC   expressing recombinant proteins for analysis and are also useful as
CC   chromosome markers or tags to identify chromosomes or to map related gene
CC   positions. The proteins are useful as amino acid supplement, carbon
CC   source, nitrogen source and carbohydrate source. The proteins are useful
CC   for treating various immune deficiencies and disorders (e.g. severe
CC   combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC   sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic
CC   reactions (e.g. asthma), myeloid or lymphoid cell deficiencies,
CC   osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g.
CC   peripheral neuropathy, Alzheimer's disease, Parkinson's disease),
CC   coagulation disorders, inflammatory diseases (e.g. systemic inflammatory
CC   response syndrome (SIRS), ischaemia-reperfusion injury, Crohn's disease),
CC   anaphylaxis and hypersensitivity. Proteins are also useful for inducing
CC   tumour immunity, for inducing bone, cartilage, tendon, ligament and/or
CC   nerve growth or regeneration, for proliferating neural cells and for
CC   regenerating nerve and brain tissue, for inducing fertility and for
CC   inhibiting tumour growth. Proteins are also useful as chemokine for
CC   mammalian cells (e.g., monocytes, fibroblasts, neutrophils), and also
CC   useful as inhibitors of receptor/ligand interactions. The present
CC   sequence represents the amino acid sequence of a human secreted protein.
XX
```

```
SQ   Sequence 538 AA;

  Query Match             88.0%;  Score 2319;  DB 7;  Length 538;
  Best Local Similarity   88.6%;  Pred. No. 1.1e-200;
  Matches  436;  Conservative   18;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:|||||||||||||||||| |||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:|| .||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492

RESULT 12
AAB75138
ID   AAB75138 standard; protein; 538 AA.
XX
AC   AAB75138;
XX
DT   08-AUG-2001  (first entry)
XX
DE   HERV-W envelope protein G.
XX
KW   Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW   envelope protein; multiple sclerosis-related superantigen; vaccine;
KW   surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW   antisense-therapy; autoimmune disorder.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   Region          271. .317
FT                   /label= SU
FT                   /note= "surface protein"
FT   Cleavage-site   317. .318
FT                   /note= "SU-TM putative cleavage site"
FT   Region          318. .538
FT                   /label= TM
FT                   /note= "transmembrane subunit"
XX
PN   WO200131021-A1.
XX
PD   03-MAY-2001.
XX
PF   30-OCT-2000; 2000WO-EP010659.
XX
PR   28-OCT-1999;   99EP-00402690.
XX
PA   (UYGE-) UNIV GENEVE.
XX
PI   Conrad B,  Mach B;
XX
DR   WPI; 2001-316336/33.
DR   N-PSDB; AAH20070.
XX
PT   New human retrovirus HERV-W ENV proteins/peptides having superantigen
PT   activity useful for diagnosing and treating multiple sclerosis.
XX
PS   Claim 1; Fig 7; 94pp; English.
XX
CC   On the basis of the PBS t-RNA motif used for the classification of human
CC   endogenous retrovirus (HERVs) the full length endogenous provirus which
CC   was been located on the long arm of human chromosome 7 (7q21-22) has been
CC   designated HERV-W. The present invention describes proteins or peptides
CC   (I) having superantigen (SAg) activity comprising the ENV protein (ENV)
CC   of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC   have neuroprotective activity, and can be used in: vaccines; antisense-
CC   therapy; and HERV-W SAg activity-inhibitors. (I) and encoding DNA/RNA are
CC   useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC   disorders. (I) are also useful for identifying substances (and optionally
CC   recovering) capable of binding to a retroviral superantigen associated
CC   with MS, substances capable of blocking SAg activity and substances
CC   capable of blocking transcription or translation of HERV-W retroviral
CC   superantigen. A protein or peptide derived from (I), modified to be
CC   devoid of SAg activity and being capable of generating an immune response
CC   against HERV-W retroviral SAg is useful in therapy. Nucleic acid
CC   molecules encoding (I) are useful as vaccines against MS. Substances
CC   capable of blocking SAg activity, capable of binding to a retroviral
CC   superantigen associated with MS, or capable of blocking transcription or
CC   translation of HERV-W retroviral superantigen for use in treating or
CC   preventing MS, obtained using (I) are useful for the treatment and
CC   prevention of MS. (I) and nucleic acids encoding them are useful for
CC   diagnosing autoimmune disease. The present sequence represents the
CC   specifically claimed envelope protein of HERV-W designated G
XX
SQ   Sequence 538 AA;

  Query Match             87.9%;  Score 2314;  DB 4;  Length 538;
  Best Local Similarity   88.4%;  Pred. No. 3.2e-200;
  Matches  435;  Conservative   19;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:|||||||||||||||||| |||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
```

```
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492


RESULT 13
AAB67652
ID   AAB67652 standard; protein; 538 AA.
XX
AC   AAB67652;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Amino acid sequence of a human endogenous retrovirus envelope protein.
XX
KW   Envelope protein; HERV; syncytia formation; placental development;
KW   syncytia; cancer; cell adhesion.
XX
OS   Human endogenous retrovirus.
XX
PN   WO200116171-A1.
XX
PD   08-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-FR002429.
XX
PR   01-SEP-1999;   99FR-00011141.
PR   15-SEP-1999;   99FR-00011793.
XX
PA   (INMR ) BIO MERIEUX.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI   Mallet F,  Cosset F,  Blond J,  Lavillette D,  Bouton O,  Ruggieri A;
XX
DR   WPI; 2001-226676/23.
DR   N-PSDB; AAF55630.
XX
PT   Detecting expression of human endogenous retrovirus envelope protein in
PT   cells of a tissue or culture, from its ability to induce syncytia.
XX
PS   Claim 1; Page 41-44; 57pp; French.
XX
CC   The present sequence represents a human endogenous retrovirus envelope
CC   protein. The specification describes a method for detecting expression of
CC   an envelope protein from a human endogenous retrovirus (HERV), in cells,
CC   of a tissue or culture. The method comprises detecting syncytia formation
CC   due to the fusogenic properties of the envelope protein. Envelope
CC   polypeptides and polynucleotides are used to produce therapeutic or
CC   prophylactic compositions, particularly for treatment of cancer, to
CC   correct defects in placental development (or other natural formation of
CC   other types of syncytia), and to promote adhesion of cells in grafts or
CC   cellular repair processes. Expression of sequences antisense to the
CC   polynucleotide are used to prevent formation of syncytia
XX
SQ   Sequence 538 AA;

  Query Match             87.9%;  Score 2314;  DB 4;  Length 538;
  Best Local Similarity   88.4%;  Pred. No. 3.2e-200;
  Matches  435;  Conservative   19;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
```

```
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492


RESULT 14
ABP40980
ID   ABP40980 standard; protein; 540 AA.
XX
AC   ABP40980;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral HERV-7q env peptide #3.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 22; Fig 4; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 540 AA;

  Query Match             87.9%;  Score 2314;  DB 3;  Length 540;
  Best Local Similarity   88.4%;  Pred. No. 3.2e-200;
  Matches  435;  Conservative   19;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | | |||||||||| | | : |||||||||||||||
Db          3 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 62

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         63 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 122

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        123 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 182

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        183 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 242

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        243 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 302

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        303 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 362

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        363 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 422

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        423 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 482

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        483 EAVKLQMEPKMQ 494


RESULT 15
ABP40988
ID   ABP40988 standard; protein; 685 AA.
XX
AC   ABP40988;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral env ORF protein.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
```

```
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 22; Page 152-154; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 685 AA;

  Query Match             87.9%;  Score 2314;  DB 3;  Length 685;
  Best Local Similarity   88.4%;  Pred. No. 4.5e-200;
  Matches  435;  Conservative   19;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | | |||||||||| | | : |||||||||||||||
Db        131 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 190

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db        191 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 250

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        251 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 310

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        311 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 370

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        371 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 430

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        431 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 490

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        491 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 550

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        551 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 610

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        611 EAVKLQMEPKMQ 622


RESULT 16
ABP41050
ID   ABP41050 standard; protein; 846 AA.
XX
```

```
AC   ABP41050;
XX
DT   01-AUG-2002  (first entry)
XX
DE   HERV-7q env protein #1.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Example 4; Fig 19; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 846 AA;

  Query Match             87.9%;  Score 2314;  DB 3;  Length 846;
  Best Local Similarity   88.4%;  Pred. No. 6.1e-200;
  Matches  435;  Conservative   19;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : ||||||||||||||| .
Db        291 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 350

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db        351 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 410

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        411 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 470

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        471 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 530

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        531 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 590

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
```

```
Db        591 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 650

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        651 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 710

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        711 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 770

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        771 EAVKLQMEPKMQ 782


RESULT 17
ADE09243
ID   ADE09243 standard; protein; 544 AA.
XX
AC   ADE09243;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Novel protein-related contig polypeptide sequence #309.
XX
KW   novel gene; novel protein; tissue marker; molecular weight marker;
KW   chromosome marker; genetic disorder; contig.
XX
OS   Unidentified.
XX
PN   WO2003054152-A2.
XX
PD   03-JUL-2003.
XX
PF   10-DEC-2002; 2002WO-US039555.
XX
PR   10-DEC-2001; 2001US-0339739P.
PR   11-DEC-2001; 2001US-0339453P.
PR   14-MAR-2002; 2002US-0365091P.
PR   14-MAR-2002; 2002US-0365384P.
PR   12-APR-2002; 2002US-0372381P.
PR   12-APR-2002; 2002US-0372615P.
PR   22-APR-2002; 2002US-00128558.
PR   24-APR-2002; 2002US-0376045P.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Asundi V,  Goodrich RW,  Ren F,  Zhang J,  Zhao QA,  Wang J;
PI   Ghosh M,  Xue AJ,  Wehrman T,  Weng G,  Zhou P,  Drmanac RT,  Wang Z;
PI   Ma Y,  Wang D,  Chen R,  Xu C,  Boyle BJ;
XX
DR   WPI; 2003-569235/53.
XX
PT   New polynucleotides, useful for expressing recombinant proteins for
PT   analysis, characterization or therapeutic use, or as markers for tissues
PT   in which the corresponding protein is preferentially expressed.
XX
PS   Disclosure; SEQ ID NO 2787; 1177pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of novel
CC   proteins. The DNA and protein sequences of the invention are useful as:
CC   markers for tissues in which the corresponding protein is preferentially
CC   expressed; as molecular weight markers on gels; as chromosome markers or
CC   tags; to identify chromosomes or to map related gene positions; and to
CC   compare with endogenous DNA sequences in patients to identify potential
CC   genetic disorders. The present amino acid sequence was used in the
CC   exemplification of the invention.
XX
SQ   Sequence 544 AA;

  Query Match             87.6%;  Score 2308.5;  DB 7;  Length 544;
  Best Local Similarity   88.6%;  Pred. No. 1e-199;
  Matches  437;  Conservative   19;  Mismatches   36;  Indels    1;  Gaps    1;
```

```
Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          6 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 65

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| ||| |||| | ||||||:
Db         66 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCRTYFTQTGMSDGGGV 125

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        126 QDQAREKHVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 185

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        186 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 245

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        246 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 305

Qy        301 TEQDLYNHVVP-KPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDME 359
              ||||||::|:   | ||||||||||| ||||| |||||| ||||||||:|||||:|||||
Db        306 TEQDLYSYVISXSPRNKRVPILPFVIGAGVLGGLGTGIGGITTSTQFYHKLSQELNGDME 365

Qy        360 QVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVK 419
              || |||||||||||||||||||||||||||||:|||||| |||| ||||||| |||||||
Db        366 QVADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLLLGEECCYYVNQSGIVTEKVK 425

Qy        420 EIRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSR 479
              ||||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| |||||
Db        426 EIRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSR 485

Qy        480 IEAVKLQMVLQME 492
              ||||||||  :|:
Db        486 IEAVKLQMEPKMQ 498


RESULT 18
ADF60170
ID   ADF60170 standard; protein; 544 AA.
XX
AC   ADF60170;
XX
DT   12-FEB-2004  (first entry)
XX
DE   Human contig polypeptide sequence SEQ ID NO:2537.
XX
KW   biological activity; genetic engineering; hybridisation probe; oligomer;
KW   primer; chromosome mapping; gene mapping; recombinant protein production;
KW   human.
XX
OS   Homo sapiens.
XX
PN   WO2003080795-A2.
XX
PD   02-OCT-2003.
XX
PF   09-AUG-2002; 2002WO-US025485.
XX
PR   09-AUG-2001; 2001US-0311261P.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Yang Y,  Wang Z,  Weng G,  Ma Y;
XX
DR   WPI; 2003-876918/81.
DR   N-PSDB; ADF59718.
XX
PT   New polynucleotides, useful as hybridization probes, oligomers or
PT   primers, for chromosome or gene mapping, for the recombinant production
PT   of proteins, and for generating antisense DNA or RNA.
XX
```

```
PS   Example 3; SEQ ID NO 2537; 571pp; English.
XX
CC   The present invention describes isolated polynucleotide sequences (I),
CC   which encode polypeptides (II) with biological activity. Also described:
CC   (1) a vector comprising (I); (2) an expression vector comprising (I); (3)
CC   a host cell genetically engineered to comprise (I) which is operatively
CC   associated with a regulatory sequence that modulates expression of (I) in
CC   the host cell; (4) a polypeptide (II) encoded by (I); (5) a composition
CC   comprising the polypeptide of (4) and a carrier; (6) an antibody directed
CC   against the polypeptide of (4); (7) detecting (I) or the polypeptide of
CC   (4) in a sample; (8) identifying a compound that binds to the polypeptide
CC   of (4); (9) producing the polypeptide of (4); and (10) a collection of
CC   polynucleotides comprising at least one of the polynucleotide sequences
CC   (I). The polynucleotides (I) can be used as hybridisation probes,
CC   oligomers or primers, for chromosome or gene mapping, for the recombinant
CC   production of proteins, and for generating antisense DNA or RNA. The
CC   present sequence represents a human contig polypeptide sequence, which is
CC   used in an example from the present invention.
XX
SQ   Sequence 544 AA;

  Query Match             87.6%;  Score 2308.5;  DB 7;  Length 544;
  Best Local Similarity   88.6%;  Pred. No. 1e-199;
  Matches  437;  Conservative   19;  Mismatches   36;  Indels    1;  Gaps    1;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          6 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 65

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| ||| |||| | ||||||:
Db         66 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCRTYFTQTGMSDGGGV 125

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        126 QDQAREKHVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 185

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        186 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 245

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        246 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 305

Qy        301 TEQDLYNHVVP-KPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDME 359
              ||||||::|:   | ||||||||||| ||||| |||||| ||||||||:|||||:|||||
Db        306 TEQDLYSYVISXSPRNKRVPILPFVIGAGVLGGLGTGIGGITTSTQFYHKLSQELNGDME 365

Qy        360 QVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVK 419
              || |||||||||||||||||||||||||||||:|||||| |||| ||||||| |||||||
Db        366 QVADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLLLGEECCYYVNQSGIVTEKVK 425

Qy        420 EIRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSR 479
              ||||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| |||||
Db        426 EIRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSR 485

Qy        480 IEAVKLQMVLQME 492
              ||||||||  :|:
Db        486 IEAVKLQMEPKMQ 498


RESULT 19
AAW99345
ID   AAW99345 standard; protein; 538 AA.
XX
AC   AAW99345;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Human endogenous retrovirus W ORF1.
XX
KW   Clone; human endogenous retrovirus; genome; autoimmune disease;
KW   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW   disseminated lupus erythematosus; pregnancy; chromosomal marker.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9902696-A1.
XX
PD   21-JAN-1999.
XX
PF   06-JUL-1998;   98WO-FR001442.
XX
PR   07-JUL-1997;   97FR-00008815.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Beseme F,  Blond J,  Bouton O,  Mandrand B,  Mallet F;
XX
DR   WPI; 1999-120897/10.
XX
PT   New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT   exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT   disease, and abnormal or failed pregnancy.
XX
PS   Example 8; Fig 6; 106pp; French.
XX
CC   This sequence represents the open reading frame 1 encoding the env
CC   protein from the human endogenous retrovirus (HERV) W (AAX25665). Nucleic
CC   acids, their fragments or peptides encoded by them derived from the HERV-
CC   W genome are markers of autoimmune disease (e.g. multiple sclerosis,
CC   rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC   dependent diabetes and related pathologies) and of abnormal or
CC   unsuccessful pregnancy and can be used as chromosomal markers for
CC   susceptibility to these conditions, or proximity markers of genes
CC   associated with this susceptibility
XX
SQ   Sequence 538 AA;

  Query Match             84.9%;  Score 2237;  DB 2;  Length 538;
  Best Local Similarity   85.6%;  Pred. No. 3e-193;
  Matches  421;  Conservative   26;  Mismatches   45;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              | |||| || :|| | | ||||||| | |||||: ||| | : |||||||||||||||
Db          1 MGLPYHIFLCSVLSPCFTLTAPPPCRCMTSSSPHPEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||  | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTGVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||:|||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNVEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | ||||||||||||||||||:||||||||||:||||||||||| ||||||| ||||||:|
Db        361 VADSLVTLQDQLNSLAAVVLRNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVEE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              | ||||  ||||:||  |||||:||||:|||||||||:|||||||||||:||| ||||||
Db        421 IPDRIQRIAEELRNTGPWGLLSRWMPWILPFLGPLAAIILLLLFGPCIFDLLVNFVSSRI 480
```

```
Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492

RESULT 20
AAM18008
ID   AAM18008 standard; protein; 263 AA.
XX
AC   AAM18008;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #4442 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 27; SEQ ID NO 22834; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 4;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179
```

```
Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263


RESULT 21
ABB37043
ID   ABB37043 standard; peptide; 263 AA.
XX
AC   ABB37043;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #4549 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 27; SEQ ID NO 29678; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a peptide encoded by a single exon nucleic acid probe
CC   of the invention. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 4;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263


RESULT 22
AAM30519
ID   AAM30519 standard; protein; 263 AA.
XX
AC   AAM30519;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #4556 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000663.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human placenta.
XX
PS   Claim 27; SEQ ID NO 30788; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders
XX
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 4;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119
```

```
Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263


RESULT 23
ABB31812
ID   ABB31812 standard; peptide; 263 AA.
XX
AC   ABB31812;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Peptide #4463 encoded by breast cell single exon nucleic acid probe.
XX
KW   Human; microarray; single exon probe; gene expression; breast; disease;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157271-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000662.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-496933/54.
XX
PT   New spatially-addressable set of single exon nucleic acid probes, useful
PT   for measuring gene expression in sample derived from human breast,
PT   comprises number of single exon nucleic acid probes.
XX
PS   Claim 27; SEQ ID NO 14780; 327pp + Sequence Listing; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human breast and BT 474 cells. The method involves contacting the
CC   probes with a collection of detectably labelled nucleic acids derived
CC   from mRNA of human breast, and then measuring the label bound to each
CC   probe of the microarray. The probes are useful for verifying the
CC   expression of regions of genomic DNA predicted to encode proteins. They
CC   are useful for gene discovery, and for determining predisposition and/or
CC   prognosing breast disease. Gene expression analysis is useful for
CC   assessing the toxicity of chemical agents on cells. The microarray of
CC   this invention presents a far greater diversity of probes for measuring
CC   gene expression, with far less bias than expressed sequence tag
CC   microarrays. The method is suitable for rapid production of functional
CC   information from genomic sequence. The present sequence is a peptide
CC   encoded by a single exon nucleic acid probe of the invention. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
```

```
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 4;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263


RESULT 24
ABB22358
ID   ABB22358 standard; protein; 263 AA.
XX
AC   ABB22358;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Protein #4357 encoded by probe for measuring heart cell gene expression.
XX
KW   Human; gene expression; heart; microarray; vascular system;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000666.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts.
XX
PS   Claim 15; SEQ ID NO 24128; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart (see
CC   ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC   probe. The probes may be used for predicting, measuring and displaying
CC   gene expression in samples derived from the human heart via microarrays.
CC   By measuring gene expression, the probes are useful for predicting,
CC   diagnosing, grading, staging, monitoring and prognosing diseases of the
CC   human heart and vascular system e.g. cardiovascular disease,
CC   hypertension, cardiac arrhythmias and congenital heart disease. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 4;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263

RESULT 25
AAM70185
ID   AAM70185 standard; protein; 263 AA.
XX
AC   AAM70185;
XX
DT   06-NOV-2001 ' (first entry)
XX
DE   Human bone marrow expressed probe encoded protein SEQ ID NO: 30491.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000668.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human bone marrow.
XX
```

```
PS   Example 4; SEQ ID NO 30491; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention
XX
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 4;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263

RESULT 26
AAM57770
ID   AAM57770 standard; protein; 263 AA.
XX
AC   AAM57770;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe encoded protein SEQ ID NO: 29875.
XX
KW   Human; brain expressed exon; gene expression analysis; probe; microarray;
KW   Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000667.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains.
```

```
XX
PS   Example 4; SEQ ID NO 29875; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is a protein encoded by one of
CC   the probes of the invention
XX
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 4;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263


RESULT 27
ABG51885
ID   ABG51885 standard; peptide; 263 AA.
XX
AC   ABG51885;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human liver peptide, SEQ ID No 30533.
XX
KW   Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW   hypercholesterolaemia; coronary heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000664.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488898/53.
XX
```

```
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human adult liver.
XX
PS   Claim 27; SEQ ID NO 30533; 658pp; English.
XX
CC   The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC   measuring human gene expression in a sample derived from human adult
CC   liver, comprising one of 13109 defined nucleotide sequences given in the
CC   specification (or complements/ fragments). The probe hybridises at high
CC   stringency to a nucleic acid molecule expressed in the human adult liver.
CC   (I) may be used for predicting, measuring and displaying gene expression
CC   in samples derived from human adult liver. The genes identified may be
CC   involved in genetic liver diseases such as cirrhosis,
CC   hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC   associated with coronary heart disease. ABG47348-ABG59930 represent human
CC   liver single exon encoded peptides of the invention. Note: The sequence
CC   information for this patent does not appear in the printed specification
CC   but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 4;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263


RESULT 28
AAM05648
ID   AAM05648 standard; protein; 263 AA.
XX
AC   AAM05648;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Peptide #4330 encoded by probe for measuring breast gene expression.
XX
KW   Probe; human; breast disease; breast cancer; development disorder;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US000661.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression in
PT   a human breast.
XX
PS   Claim 27; SEQ ID NO 14388; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes
CC   (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for measuring human gene expression in
CC   a human breast sample, where the probe hybridises at high stringency to a
CC   nucleic acid expressed in the human breast. The probes are useful for
CC   predicting, diagnosing, grading, staging, monitoring and prognosing
CC   diseases of the human breast, particularly those diseases with polygenic
CC   aetiology. The diseases include: breast cancer, disorders of development,
CC   inflammatory diseases of the breast, fibrocystic changes, proliferative
CC   breast disease and non-carcinoma tumours. Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 4;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263


RESULT 29
ABG39819
ID   ABG39819 standard; peptide; 263 AA.
XX
AC   ABG39819;
XX
DT   19-AUG-2002  (first entry)
XX
DE   Human peptide encoded by genome-derived single exon probe SEQ ID 29484.
XX
KW   Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
```

```
KW   hyaline membrane disease.
XX
OS   Homo sapiens.
XX
PN   WO200186003-A2.
XX
PD   15-NOV-2001.
XX
PF   30-JAN-2001; 2001WO-US000665.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2002-114183/15.
XX
PT   Spatially-addressable set of single exon nucleic acid probes, used to
PT   measure gene expression in human lung samples.
XX
PS   Claim 27; SEQ ID NO 29484; 634pp; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human lung comprising single exon nucleic acid probes having one of
CC   12614 nucleic acid sequences mentioned in the specification, or their
CC   complements or the 12387 open reading frames derived from the 12614
CC   probes. Also included are a microarray comprising the novel set of probes
CC   ; the novel set of probes which hybridise at high stringency to a nucleic
CC   acid expressed in the human lung; measuring gene expression in a sample
CC   derived from human lung, comprising (a) contacting the array with a
CC   collection of detectably labeled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of the
CC   array; identifying exons in a eukaryotic genome, comprising (a)
CC   algorithmically predicting at least one exon from genomic sequences of
CC   the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12011 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene expression
CC   analysis, and for identifying exons in a gene, particularly using human
CC   lung derived mRNA and for the study of lung diseases such as asthma, lung
CC   cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC   disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC   tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC   Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC   histiocytosis, lymphangioleiomyomtosis, pulmonary alveolar proteinosis,
CC   Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC   dyskinesis, pulmonary hypertension and hyaline membrane disease. The
CC   present sequence is a peptide/protein encoded by a single exon probe of
CC   the invention. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 263 AA;

  Query Match             46.9%;  Score 1235.5;  DB 5;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 5.1e-103;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;
```

```
Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263


RESULT 30
ABG07532
ID   ABG07532 standard; protein; 446 AA.
XX
AC   ABG07532;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #7523.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS71719.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 37891; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
```

```
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 446 AA;

  Query Match             42.9%;  Score 1130.5;  DB 4;  Length 446;
  Best Local Similarity   68.1%;  Pred. No. 3.5e-93;
  Matches  245;  Conservative   23;  Mismatches   41;  Indels   51;  Gaps   11;

Qy         92 INPSC--PGGLGATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLT----RGH----- 140
              | | |  | ||              ||     | | :  | : :||::     |
Db        116 IIPRCGLPRGL-------------QSD----NGPAFKATVTQGLSQMSDTSWEGRYPVIL 158

Qy        141 STPSPYK-GLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN--CWMCLPLH---- 193
              |||:  |   | |::|     |||:          |      || :   : | ||
Db        159 STPTAVKVAGVESRIH-----HTRVKPW-------LLPKEPGNPGDNASYSCEPLEDLGL 206

Qy        194 -FR---PYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSS 249
               |:   |:||||||||||||||||||||:|||||||||||||||||| ||||||||||:|
Db        207 LFKQQPPHISIPVPEQWNNFSTEINTTSILVGPLVSNLEITHTSNLTRVKFSNTIDTTNS 266

Qy        250 QCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHV 309
              ||||||||||:|||||||||||||||||||||||| ||||||||||| ||||||||||||:|
Db        267 QCIRWVTPPTQIVCLPSGIFFVCGTSAYHCLNGFSESMCFLSFLVHPMTIYTEQDLYNYV 326

Qy        310 VPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQ 369
              |||| ||||||||||||||:|||||||:||||||||||||||||:|||||:| |||||||
Db        327 VPKPRNKRVPILPFVIRAGMLGRLGTGMGSITTSTQFYYKLSQELNGDMERVADSLVTLQ 386

Qy        370 DQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRA 429
              ||||||||||||||||||||||:||||||||||| ||||||| |:||||||||| || ||
Db        387 DQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIITEKVKEIRDGIQRRA 446


Search completed: January 22, 2005, 22:09:22
Job time : 122 secs
```

```
                  GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.


OM protein - protein search, using sw model

Run on:         January 22, 2005, 21:09:06 ; Search time 45 Seconds
                                          (without alignments)
                                          726.551 Million cell updates/sec

Title:          US-09-319-156B-10
Perfect score:  2634
Sequence:       1 MALPYHTFLFTVLLPPFALT..........KFVSSRIEAVKLQMVLQMEP 493

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters:      478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                     %
Result             Query
  No.     Score    Match Length DB  ID                           Description
-----------------------------------------------------------------------------
     1     2632     99.9    493  4  US-08-979-847B-106           Sequence 106, App
     2     2319     88.0    538  3  US-09-175-928-4              Sequence 4, Appli
     3    559.5     21.2    567  1  US-08-007-282B-2             Sequence 2, Appli
     4      516     19.6    162  4  US-08-979-847B-109           Sequence 109, App
     5      452     17.2    433  3  US-08-691-563C-87            Sequence 87, Appl
     6      452     17.2    433  4  US-09-374-766-87             Sequence 87, Appl
     7      452     17.2    433  4  US-08-979-847B-81            Sequence 81, Appl
     8    407.5     15.5    632  4  US-09-315-127-2              Sequence 2, Appli
     9    407.5     15.5    632  4  US-09-315-127-3              Sequence 3, Appli
    10    407.5     15.5    665  4  US-09-309-572-14             Sequence 14, Appl
    11    407.5     15.5    665  4  US-09-718-096-14             Sequence 14, Appl
    12    407.5     15.5   1312  4  US-09-554-572-26             Sequence 26, Appl
    13    389.5     14.8    673  3  US-09-075-272-5              Sequence 5, Appli
    14      385     14.6    654  4  US-09-315-127-11             Sequence 11, Appl
    15      385     14.6    654  4  US-09-315-127-12             Sequence 12, Appl
    16    376.5     14.3    645  4  US-09-315-127-8              Sequence 8, Appli
    17    376.5     14.3    645  4  US-09-315-127-9              Sequence 9, Appli
    18    376.5     14.3    660  3  US-09-111-085-2              Sequence 2, Appli
    19    376.5     14.3    660  3  US-09-376-781-5              Sequence 5, Appli
    20      368     14.0    638  3  US-09-376-781-6              Sequence 6, Appli
    21      364     13.8    192  3  US-08-486-099-107            Sequence 107, App
    22      364     13.8    192  3  US-08-360-107A-117           Sequence 117, App
    23      364     13.8    192  3  US-08-484-223B-107           Sequence 107, App
    24      364     13.8    192  3  US-08-919-597-107            Sequence 107, App
    25      364     13.8    192  3  US-08-475-668A-107           Sequence 107, App
    26      364     13.8    192  3  US-08-485-551A-107           Sequence 107, App
    27      364     13.8    192  3  US-08-471-913A-107           Sequence 107, App
    28      364     13.8    192  3  US-08-485-264A-107           Sequence 107, App
    29      364     13.8    192  3  US-08-474-349A-107           Sequence 107, App
    30      364     13.8    192  4  US-08-470-896-107            Sequence 107, App
    31      364     13.8    192  4  US-08-485-546A-107           Sequence 107, App
    32    357.5     13.6    667  4  US-09-315-127-5              Sequence 5, Appli
    33    357.5     13.6    667  4  US-09-315-127-6              Sequence 6, Appli
    34    357.5     13.6    667  4  US-09-070-630-13             Sequence 13, Appl
    35      352     13.4    232  4  US-09-555-352-10             Sequence 10, Appl
    36      349     13.2    657  3  US-09-111-085-4              Sequence 4, Appli
    37      349     13.2    657  3  US-09-376-781-4              Sequence 4, Appli
    38      347     13.2    657  4  US-09-171-553B-10            Sequence 10, Appl
    39      345     13.1    656  3  US-09-376-781-3              Sequence 3, Appli
    40      345     13.1    656  4  US-09-171-553B-6             Sequence 6, Appli
    41      341     12.9    191  4  US-09-350-841A-1593          Sequence 1593, Ap
    42      311     11.8    196  2  US-08-484-126-7              Sequence 7, Appli
    43      311     11.8    196  4  US-09-374-909-7              Sequence 7, Appli
    44      298     11.3     67  1  US-08-686-878A-49            Sequence 49, Appl
    45      298     11.3     67  1  US-08-721-489-3              Sequence 3, Appli


                                   ALIGNMENTS


RESULT 1
US-08-979-847B-106
; Sequence 106, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,847B
;            FILING DATE: 26-No. 6582703-1997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;  INFORMATION FOR SEQ ID NO: 106:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 493 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-08-979-847B-106
```

```
  Query Match             99.9%;  Score 2632;  DB 4;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-262;
  Matches  493;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              |||||||||||||
Db        481 EAVKLQMVLQMEP 493


RESULT 2
US-09-175-928-4
; Sequence 4, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT: McCoy, John M.
;  APPLICANT: LaVallie, Edward R.
;  APPLICANT: Collins-Racie, Lisa A.
;  APPLICANT: Evans, Cheryl
;  APPLICANT: Merberg, David
;  APPLICANT: Treacy, Maurice
;  APPLICANT: Mi, Sha
;  APPLICANT: Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6006B.AJ172A
;  CURRENT APPLICATION NUMBER: US/09/175,928A
;  CURRENT FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-175-928-4

  Query Match             88.0%;  Score 2319;  DB 3;  Length 538;
  Best Local Similarity   88.6%;  Pred. No. 5.9e-230;
  Matches  436;  Conservative   18;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
```

```
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492


RESULT 3
US-08-007-282B-2
; Sequence 2, Application US/08007282B
; Patent No. 5403582
;  GENERAL INFORMATION:
;    APPLICANT:  NAZERIAN, KEYVAN
;    APPLICANT:  CALVERT, JAY G.
;    APPLICANT:  WITTER, RICHARD L.
;    APPLICANT:  YANAGIDA, NOBORU
;    TITLE OF INVENTION:  VACCINE COMPRISING FOWLPOX VIRUS
;    TITLE OF INVENTION:  RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF AN
;    TITLE OF INVENTION:  AVIAN RETICULOENDOTHELIOSIS RETROVIRUS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  8110 Gatehouse Road  Suite 500 East
;      CITY:  Falls Church
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/007,282B
;      FILING DATE:  19930121
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY JR., GERALD M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1644-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  567 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-007-282B-2

  Query Match             21.2%;  Score 559.5;  DB 1;  Length 567;
  Best Local Similarity   31.9%;  Pred. No. 1.5e-48;
  Matches  153;  Conservative   75;  Mismatches  158;  Indels   93;  Gaps   19;

Qy         66 THMPRNCYNSATLCMHANTHYWT---------------GKMINPSCPGGLGATVCWTYFT 110
              | |   ||     |      |:|                |:|  || | :|  |||
Db        121 TQMHSTCYEKTQECTLLGKTYFTAILQKTKLGSYEDGPNKLIQASCTGTVGKPVCWDPVA 180

Qy        111 HTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPS-PYKGLVLSKLHETLRTHTRLVSLFN 169
                 :|||||     ||: |:| : :: | || ||  |  | | :        :| | |
Db        181 PVYVSDGGGPTDMIREESVRERLEEIIR-HSYPSVQYHPLALPR--------SRGVDLDP 231

Qy        170 TT---LTRLHEV-SAQNP---TNCWMCLPLHFRPYISIPVPEQWN-----NFSTEI---- 213
               |   |   |:| :| ||    |||:|: |     |   :|   |     | |  :
Db        232 QTSDILEATHQVLNATNPKLAENCWLCMTL--GTPIPAAIPTNGNVTLDGNCSLSLPFGC 289

Qy        214 ----------------NTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTP 257
                              | | : ||         | :| | ::    :   :||    :
Db        290 NPPGSIDVSCYAGEADNRTGIPVG-------YVHFTNCTSIQ---EVTNETSQ----MGN 335

Qy        258 PTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP--- 313
               ||: | | |  |||| : ||  |      :| |: :|| ::| : ::     :| |
Db        336 LTRL-CPPPGHVFVCGNNMAYTALPNKWIGLCILASIVPDISIISGEE----PIPLPSIE 390

Qy        314 -----HNKRVPILPFVIRAGVLGRL---GTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
                   | : |  :| ::  |: |     |||:|    |   |:||| ::  |:: :: ::
Db        391 YTARRHKRAVQFIPLLVGLGISGATLAGGTGLG---VSVHTYHKLSNQLIEDVQALSGTI 447

Qy        366 VTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI 425
                ||||::||| ||||||| ||||||::|| || | |: |:| |:| || :|:::::: :
Db        448 NDLQDQIDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIRKLQEDL 507

Qy        426 QCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVK 484
                |   | :   |  |: ::|::|| ||||  ||| |  ||||   | : :  :|:|||
Db        508 IERKRALYDNPLWSGLNGFLPYLLPLLGPLFGLILFLTLGPCIMKTLTRIIHDKIQAVK 566

RESULT 4
US-08-979-847B-109
; Sequence 109, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;     INFORMATION FOR SEQ ID NO: 109:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 162 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 109:
US-08-979-847B-109

  Query Match             19.6%;  Score 516;  DB 4;  Length 162;
  Best Local Similarity   87.6%;  Pred. No. 6.4e-45;
  Matches   99;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy        381 QNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGL 440
              ||||||||||||||||||||||| | ||||| |:||||||| ||| ||||:||||  |||
Db          1 QNRRALDLLTAKRGGTCLFLGEECCXYVNQSGIITEKVKEIXDRIXCRAEDLQNTAPWGL 60

Qy        441 LSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
              ||||||| |||||||||:| |||||||||| |||||||||||||||:||||||
Db         61 LSQWMPWTLPFLGPLAAIIFLLLFGPCIFNFLVKFVSSRIEAVKLQIVLQMEP 113


RESULT 5
US-08-691-563C-87
; Sequence 87, Application US/08691563C
; Patent No. 6001987
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTI
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563C
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  87:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  433 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-691-563C-87

  Query Match             17.2%;  Score 452;  DB 3;  Length 433;
  Best Local Similarity   54.3%;  Pred. No. 1.2e-37;
  Matches   89;  Conservative   27;  Mismatches   38;  Indels   10;  Gaps    2;

Qy        334 GTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||||  ::||  :|: ||:  :  ::::  |::||| ||:||||: |||||   ||||::
Db          6 GTGIAGLSTSLSYYHTLSKNFSDSLQEIMKSILTLQSQLDSLAAMTLQNRRGPHLLTAEK 65

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI----QCRAEELQNTERWGLLSQWMPWVL 449
              || | ||||| |:| ||| || :    :::|     || :    |      |  |  |:|
Db         66 GGLCTFLGEECCFYTNQSGIVRDATWHLQERASDIRQCLSNSYTN------LWSWATWLL 119

Qy        450 PFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
              |||||:||::||| |||||| ||||||||||||:||||||||||
Db        120 PFLGPMAAILLLLTFGPCIFKLLVKFVSSRIEAIKLQMVLQMEP 163


RESULT 6
US-09-374-766-87
; Sequence 87, Application US/09374766
; Patent No. 6579526
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/374,766
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  87:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  433 amino acids
```

```
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-374-766-87

  Query Match             17.2%;  Score 452;  DB 4;  Length 433;
  Best Local Similarity   54.3%;  Pred. No. 1.2e-37;
  Matches   89;  Conservative   27;  Mismatches   38;  Indels   10;  Gaps    2;

Qy        334 GTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||||  ::||  :|: ||:  :  ::::  |::||| ||:||||: |||||   ||||::
Db          6 GTGIAGLSTSLSYYHTLSKNFSDSLQEIMKSILTLQSQLDSLAAMTLQNRRGPHLLTAEK 65

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI----QCRAEELQNTERWGLLSQWMPWVL 449
              || | ||||| |:| ||| || :    :::|     || :    |      |  |  |:|
Db         66 GGLCTFLGEECCFYTNQSGIVRDATWHLQERASDIRQCLSNSYTN------LWSWATWLL 119

Qy        450 PFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
              |||||:||::||| |||||| ||||||||||||:||||||||||
Db        120 PFLGPMAAILLLLTFGPCIFKLLVKFVSSRIEAIKLQMVLQMEP 163


RESULT 7
US-08-979-847B-81
; Sequence 81, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLA
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 81:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 433 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: not relevant
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 81:
US-08-979-847B-81
```

```
  Query Match            17.2%;  Score 452;  DB 4;  Length 433;
  Best Local Similarity  54.3%;  Pred. No. 1.2e-37;
  Matches   89;  Conservative   27;  Mismatches   38;  Indels   10;  Gaps    2;

Qy        334 GTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||||  ::||  :|: ||:  :  ::::  |::||| ||:||||: |||||   ||||::
Db          6 GTGIAGLSTSLSYYHTLSKNFSDSLQEIMKSILTLQSQLDSLAAMTLQNRRGPHLLTAEK 65

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI----QCRAEELQNTERWGLLSQWMPWVL 449
              || | ||||| |:| ||| || :    :::|     || :    |      |  |  |:|
Db         66 GGLCTFLGEECCFYTNQSGIVRDATWHLQERASDIRQCLSNSYTN------LWSWATWLL 119

Qy        450 PFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
              |||||:||::||| |||||| ||||||||||||:||||||||||
Db        120 PFLGPMAAILLLLTFGPCIFKLLVKFVSSRIEAIKLQMVLQMEP 163

RESULT 8
US-09-315-127-2
; Sequence 2, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
;  APPLICANT: The University of Tennessee, c/o Richard Cox
;  TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
;  TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
;  FILE REFERENCE: 44137-5023, U. of Tennessee
;  CURRENT APPLICATION NUMBER: US/09/315,127
;  CURRENT FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 632
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-315-127-2

  Query Match            15.5%;  Score 407.5;  DB 4;  Length 632;
  Best Local Similarity  25.0%;  Pred. No. 8.5e-33;
  Matches  146;  Conservative   77;  Mismatches  227;  Indels  133;  Gaps   21;

Qy         16 PFALTAPPPCCCTTSS--------SPYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF-- 62
              ||:    ||||   ||         |      |     |   :|  :::|    |  |
Db         63 PFSSPPGPPCCSGGSSPGCSRDCEEPLTSLTPRCNTAWNRLKLDQTTHKS----NEGFYV 118

Qy         63 TAHTHMPR---NC------YNSATLCMHANTHYWTGK------MINPSCPGGLGATVC-- 105
                  | ||   :|      | :   |      ||          :| :        ||
Db        119 CPGPHRPRESKSCGGPDSFYCAYWGCETTGRAYWKPSSSWDFITVNNNLTSDQAVQVCKD 178

Qy        106 --------------------------WTYFTHTSMSDGG---GI---------------- 120
                                        |    : |  | |   ||
Db        179 NKWCNPLVIRFTDAGRRVTSWTTGHYWGLRLYVSGQDPGLTFGIRLRYQNLGPRVPIGPN 238

Qy        121 ------QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH-ETLRTHTRLVSLFNTTLT 173
                    |  :: | ||          |   |  |  ||        |  ||::| :
Db        239 PVLADQQPLSKPKPVKSP--------SVTKPPSGTPLSPTQLPPAGTENRLLNLVDGAYQ 290

Qy        174 RLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT--- 230
               |:  |      ||:||      |  : |   ::| ::     ||     ::  |:|
Db        291 ALNLTSPDKTQECWLCLVAGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQG 350

Qy        231 -------HTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGS 283
                      |    |    ||  |:|      |         |:|  : | |    |:: :
Db        351 LCIGAVPKTHQALC----NTTQTSSRGSYYLVA--------PTGTMWACSTGLTPCISTT 398

Qy        284 ----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTG 336
                  :   | |  | | :| ::   :|        :|| |:   |  ::    :|  : |
Db        399 ILNLTTDYCVLVELWPRVTYHSPSYVYGLFERSNRHKREPVSLTLALLLGGLTMGGIAAG 458

Qy        337 IGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||: ||   :|| : :|   :  |: :|  |:  |:  | ||: ||||||| ||||  |
Db        459 IGTGTTALMATQQFQQLQAAVQDDLREVEKSISNLEKSLTSLSEVVLQNRRGLDLLFLKE 518

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW---- 447
              || |  | || |:| : : :| : : ::|:|:  | :  ::|: |  || ::  ||
Db        519 GGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFESTQGWFEGLFNR-SPWFTTL 577

Qy        448 VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              :   :|||  |:::||||||| | ||:||  ||  |:  :: |
Db        578 ISTIMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALVLTQ 620

RESULT 9
US-09-315-127-3
; Sequence 3, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
;  APPLICANT: The University of Tennessee, c/o Richard Cox
;  TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
;  TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
;  FILE REFERENCE: 44137-5023, U. of Tennessee
;  CURRENT APPLICATION NUMBER: US/09/315,127
;  CURRENT FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 632
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SEQ ID NO.2,
;   OTHER INFORMATION: envelope protein produced by retroviral vector of
;   OTHER INFORMATION: seq. id no.1
US-09-315-127-3

  Query Match            15.5%;  Score 407.5;  DB 4;  Length 632;
  Best Local Similarity  25.0%;  Pred. No. 8.5e-33;
  Matches  146;  Conservative   77;  Mismatches  227;  Indels  133;  Gaps   21;

Qy         16 PFALTAPPPCCCTTSS--------SPYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF-- 62
              ||:    ||||   ||         |      |     |   :|  :::|    |  |
Db         63 PFSSPPGPPCCSGGSSPGCSRDCEEPLTSLTPRCNTAWNRLKLDQTTHKS----NEGFYV 118

Qy         63 TAHTHMPR---NC------YNSATLCMHANTHYWTGK------MINPSCPGGLGATVC-- 105
                  | ||   :|      | :   |      ||          :| :        ||
Db        119 CPGPHRPRESKSCGGPDSFYCAYWGCETTGRAYWKPSSSWDFITVNNNLTSDQAVQVCKD 178

Qy        106 --------------------------WTYFTHTSMSDGG---GI---------------- 120
                                        |    : |  | |   ||
Db        179 NKWCNPLVIRFTDAGRRVTSWTTGHYWGLRLYVSGQDPGLTFGIRLRYQNLGPRVPIGPN 238

Qy        121 ------QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH-ETLRTHTRLVSLFNTTLT 173
                    |  :: | ||          |   |  |  ||        |  ||::| :
Db        239 PVLADQQPLSKPKPVKSP--------SVTKPPSGTPLSPTQLPPAGTENRLLNLVDGAYQ 290

Qy        174 RLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT--- 230
               |:  |      ||:||      |  : |   ::| ::     ||     ::  |:|
Db        291 ALNLTSPDKTQECWLCLVAGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQG 350

Qy        231 -------HTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGS 283
                      |    |    ||  |:|      |         |:|  : | |    |:: :
Db        351 LCIGAVPKTHQALC----NTTQTSSRGSYYLVA--------PTGTMWACSTGLTPCISTT 398

Qy        284 ----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTG 336
                  :   | |  | | :| ::   :|        :|| |:   |  ::    :|  : |
Db        399 ILNLTTDYCVLVELWPRVTYHSPSYVYGLFERSNRHKREPVSLTLALLLGGLTMGGIAAG 458

Qy        337 IGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||: ||   :|| : :|   :  |: :|  |:  |:  | ||: ||||||| ||||  |
Db        459 IGTGTTALMATQQFQQLQAAVQDDLREVEKSISNLEKSLTSLSEVVLQNRRGLDLLFLKE 518

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW---- 447
              || |  | || |:| : : :| : : ::|:|:  | :  ::|: |  || ::  ||
Db        519 GGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFESTQGWFEGLFNR-SPWFTTL 577

Qy        448 VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              :   :|||  |:::||||||| | ||:||  ||  |:  :: |
Db       578 ISTIMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALVLTQ 620

RESULT 10
US-09-309-572-14
; Sequence 14, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
;  APPLICANT: Heinrich-Pette-Institut
;  TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
;  FILE REFERENCE: P50489
;  CURRENT APPLICATION NUMBER: US/09/309,572
;  CURRENT FILING DATE: 1999-05-11
;  EARLIER APPLICATION NUMBER: DE 198 56 463
;  EARLIER FILING DATE: 1998-11-26
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 665
;   TYPE: PRT
;   ORGANISM: Moloney murine leukemia virus
;   FEATURE:
;   OTHER INFORMATION: env protein
US-09-309-572-14

  Query Match             15.5%;  Score 407.5;  DB 4;  Length 665;
  Best Local Similarity   25.0%;  Pred. No. 9.2e-33;
  Matches  146;  Conservative   77;  Mismatches  227;  Indels  133;  Gaps   21;

Qy         16 PFALTAPPPCCCTTSS--------SPYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF-- 62
              ||:    ||||   ||         |      |     |   :|  :::|    |  |
Db         96 PFSSPPGPPCCSGGSSPGCSRDCEEPLTSLTPRCNTAWNRLKLDQTTHKS----NEGFYV 151

Qy         63 TAHTHMPR---NC------YNSATLCMHANTHYWTGK------MINPSCPGGLGATVC-- 105
                  | ||   :|      | :   |      ||          :| :        ||
Db        152 CPGPHRPRESKSCGGPDSFYCAYWGCETTGRAYWKPSSSWDFITVNNNLTSDQAVQVCKD 211

Qy        106 -------------------------WTYFTHTSMSDGG---GI--------------- 120
                                        |    : |  | |   ||
Db        212 NKWCNPLVIRFTDAGRRVTSWTTGHYWGLRLYVSGQDPGLTFGIRLRYQNLGPRVPIGPN 271

Qy        121 ------QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH-ETLRTHTRLVSLFNTTLT 173
                    |  :: | ||          |   |  |  ||        |  ||::| :
Db        272 PVLADQQPLSKPKPVKSP--------SVTKPPSGTPLSPTQLPPAGTENRLLNLVDGAYQ 323

Qy        174 RLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT--- 230
               |:  |      ||:||      |  : |   ::| ::     ||     ::  |:|
Db        324 ALNLTSPDKTQECWLCLVAGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQG 383

Qy        231 -------HTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGS 283
                      |    |    ||  |:|      |         |:|  : | |    |:: :
Db        384 LCIGAVPKTHQALC----NTTQTSSRGSYYLVA--------PTGTMWACSTGLTPCISTT 431

Qy        284 ----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTG 336
                  :   | |  | |:| ::   :|        :|| |:   |  ::    :| :  |
Db        432 ILNLTTDYCVLVELWPRVTYHSPSYVYGLFERSNRHKREPVSLTLALLLGGLTMGGIAAG 491

Qy        337 IGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||: ||   :|| : :|   :  |: :|  |:  |:  | ||: ||||||| ||||  |
Db        492 IGTGTTALMATQQFQQLQAAVQDDLREVEKSISNLEKSLTSLSEVVLQNRRGLDLLFLKE 551

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW---- 447
              || |  | || |:| : : :| : : ::|:|:  | :  ::|: |  || ::  ||
Db        552 GGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFESTQGWFEGLFNR-SPWFTTL 610

Qy        448 VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              :   :|||  |:::||||||| | ||:||  ||  |:  :: |
Db        611 ISTIMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALVLTQ 653

RESULT 11
US-09-718-096-14
; Sequence 14, Application US/09718096
```

```
; Patent No. 6589763
; GENERAL INFORMATION:
;  APPLICANT:  Von Laer, Meike-Dorothee
;  TITLE OF INVENTION:  RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
;  FILE REFERENCE:  35-195
;  CURRENT APPLICATION NUMBER: US/09/718,096
;  CURRENT FILING DATE:  2000-11-22
;  PRIOR APPLICATION NUMBER:  DE 19856463.5
;  PRIOR FILING DATE:  1998-11-26
;  PRIOR APPLICATION NUMBER:  EP 99250415.9
;  PRIOR FILING DATE:  1999-11-25
;  PRIOR APPLICATION NUMBER:  US 09/309,572
;  PRIOR FILING DATE:  1999-05-11
;  NUMBER OF SEQ ID NOS:  26
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 665
;   TYPE: PRT
;   ORGANISM: Moloney murine leukemia virus
;   FEATURE:
;   OTHER INFORMATION: env protein
US-09-718-096-14

  Query Match             15.5%;  Score 407.5;  DB 4;  Length 665;
  Best Local Similarity   25.0%;  Pred. No. 9.2e-33;
  Matches  146;  Conservative   77;  Mismatches  227;  Indels  133;  Gaps   21;

Qy         16 PFALTAPPPCCCTTSS--------SPYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF-- 62
              ||:    ||||   ||         |      |     |   :|  :::|    |  |
Db         96 PFSSPPGPPCCSGGSSPGCSRDCEEPLTSLTPRCNTAWNRLKLDQTTHKS----NEGFYV 151

Qy         63 TAHTHMPR---NC------YNSATLCMHANTHYWTGK------MINPSCPGGLGATVC-- 105
                  | ||   :|      | :   |      ||          :| :        ||
Db        152 CPGPHRPRESKSCGGPDSFYCAYWGCETTGRAYWKPSSSWDFITVNNNLTSDQAVQVCKD 211

Qy        106 -------------------------WTYFTHTSMSDGG---GI--------------- 120
                                        |    : |  | |   ||
Db        212 NKWCNPLVIRFTDAGRRVTSWTTGHYWGLRLYVSGQDPGLTFGIRLRYQNLGPRVPIGPN 271

Qy        121 ------QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH-ETLRTHTRLVSLFNTTLT 173
                    |  :: | ||          |   |  |  ||        |  ||::| :
Db        272 PVLADQQPLSKPKPVKSP--------SVTKPPSGTPLSPTQLPPAGTENRLLNLVDGAYQ 323

Qy        174 RLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT--- 230
               |:  |      ||:||      |  : |   ::| ::     ||     ::  |:|
Db        324 ALNLTSPDKTQECWLCLVAGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQG 383

Qy        231 -------HTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGS 283
                      |    |    ||  |:|      |         |:|  : | |    |:: :
Db        384 LCIGAVPKTHQALC----NTTQTSSRGSYYLVA--------PTGTMWACSTGLTPCISTT 431

Qy        284 ----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTG 336
                  :   | |  | |:| ::   :|        :|| |:   |  ::    :| :  |
Db        432 ILNLTTDYCVLVELWPRVTYHSPSYVYGLFERSNRHKREPVSLTLALLLGGLTMGGIAAG 491

Qy        337 IGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||: ||   :|| : :|   :  |: :|  |:  |:  | ||: ||||||| ||||  |
Db        492 IGTGTTALMATQQFQQLQAAVQDDLREVEKSISNLEKSLTSLSEVVLQNRRGLDLLFLKE 551

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW---- 447
              || |  | || |:| : : :| : : ::|:|:  | :  ::|: |  || ::  ||
Db        552 GGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFESTQGWFEGLFNR-SPWFTTL 610

Qy        448 VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              :   :|||  |:::||||||| | ||:||  ||  |:  :: |
Db        611 ISTIMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALVLTQ 653

RESULT 12
US-09-554-572-26
; Sequence 26, Application US/09554572
; Patent No. 6573091
; GENERAL INFORMATION:
```

```
;  APPLICANT: NATURE TECHNOLOGY, INC.
;  TITLE OF INVENTION: CHIMERIC VIRAL PACKAGING SIGNAL WITHOUT GAG GENE
;  TITLE OF INVENTION:  SEQUENCES
;  FILE REFERENCE: 228.00030201
;  CURRENT APPLICATION NUMBER: US/09/554,572
;  CURRENT FILING DATE:  2000-09-18
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
;   LENGTH: 1312
;   TYPE: PRT
;   ORGANISM: Murine leukemia virus
US-09-554-572-26

  Query Match             15.5%;  Score 407.5;  DB 4;  Length 1312;
  Best Local Similarity   25.0%;  Pred. No. 2.7e-32;
  Matches  146;  Conservative   77;  Mismatches  227;  Indels  133;  Gaps   21;

Qy          16 PFALTAPPPCCCTTSS--------SPYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF-- 62
               ||:    ||||   ||         |      |     |   :|  :::|    |  |
Db         743 PFSSPPGPPCCSGGSSPGCSRDCEEPLTSLTPRCNTAWNRLKLDQTTHKS----NEGFYV 798

Qy          63 TAHTHMPR---NC------YNSATLCMHANTHYWTGK------MINPSCPGGLGATVC-- 105
                   | ||   :|      | :   |      ||          :| :        ||
Db         799 CPGPHRPRESKSCGGPDSFYCAYWGCETTGRAYWKPSSSWDFITVNNNLTSDQAVQVCKD 858

Qy         106 --------------------------WTYFTHTSMSDGG---GI---------------- 120
                                         |    : |  | |   ||
Db         859 NKWCNPLVIRFTDAGRRVTSWTTGHYWGLRLYVSGQDPGLTFGIRLRYQNLGPRVPIGPN 918

Qy         121 ------QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH-ETLRTHTRLVSLFNTTLT 173
                     |  :: | ||          |   |  |  ||        |  ||::| :
Db         919 PVLADQQPLSKPKPVKSP--------SVTKPPSGTPLSPTQLPPAGTENRLLNLVDGAYQ 970

Qy         174 RLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT--- 230
                |:  |      ||:||      |  : |   ::| ::     ||     ::  |:|
Db         971 ALNLTSPDKTQECWLCLVAGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQG 1030

Qy         231 -------HTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGS 283
                       |    |    ||  |:|      |         |:|  : | |    |:: :
Db        1031 LCIGAVPKTHQALC----NTTQTSSRGSYYLVA--------PTGTMWACSTGLTPCISTT 1078

Qy         284 ----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTG 336
                   :   | |  | | :| ::   :|        :|| |:   |  ::    :| :  |
Db        1079 ILNLTTDYCVLVELWPRVTYHSPSYVYGLFERSNRHKREPVSLTLALLLGGLTMGGIAAG 1138

Qy         337 IGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
               ||: ||   :|| : :|   :  |: :|  |:  |:  | ||: ||||||| ||||  |
Db        1139 IGTGTTALMATQQFQQLQAAVQDDLREVEKSISNLEKSLTSLSEVVLQNRRGLDLLFLKE 1198

Qy         394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW---- 447
               || |  | || |:| : : :| : : ::|:|:  | :  ::|: |  || ::  ||
Db        1199 GGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFESTQGWFEGLFNR-SPWFTTL 1257

Qy         448 VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
               :   :|||  |:::||||||| | ||:||  ||  |:  :: |
Db        1258 ISTIMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALVLTQ 1300

RESULT 13
US-09-075-272-5
; Sequence 5, Application US/09075272
; Patent No. 6136598
;  GENERAL INFORMATION:
;    APPLICANT:  MILLER, A. DUSTY
;    APPLICANT:  WOLGAMOT, GREG
;    APPLICANT:  BONHAM, LYNN
;    TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
;    TITLE OF INVENTION:  PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,272
;      FILING DATE:  08-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  673 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-075-272-5

  Query Match             14.8%;  Score 389.5;  DB 3;  Length 673;
  Best Local Similarity   27.7%;  Pred. No. 6.7e-31;
  Matches  106;  Conservative   62;  Mismatches  130;  Indels   85;  Gaps   13;

Qy         158 LRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTS 217
               | |  |||||       |:  :     :||:|       | | |
Db         317 LDTENRLVSLVQGAFLVLNRTNPNMTQSCWLC-------YASNP---------------- 353

Qy         218 VLVGPLVSNLEITHTSNLT-----CV-----KFSNTIDTTSSQCI--------------- 252
                   |    :  | | |:|     |:     | : |  : :  |:
Db         354 ----PYYEGIAQTRTYNITSDHSQCLWGENRKLTLTAVSGNGLCLGQVPQDKWHLCNQTQ 409

Qy         253 --------RWVTPPTRIVCLPSGIFFVCGTSAYHCLN----GSSESMCFLSFLVPPMTIY 300
                       ::: ||   |       : | |    |::     ||:  | |  |:| :  :
Db         410 NIRPNKGGQYLVPPIDTV-------WACNTGLTPCISMSVFNSSKDFCILVQLIPRLLYH 462

Qy         301 TEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGR-LGTGIGSITTSTQFYYKLSQEING 356
                :    :    :   || ||   |  ::  ||    :|||  ::  : ::: :|   ::
Db         463 DDSSFLDKFEHRVRWKREPITLTLAVLLGLGVAAAGVGTGTAALIQTPRYFEELRTAMDT 522

Qy         357 DMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTE 416
               |:  :  |:  |:: | ||: ||||||| ||||  | || |  | || |:||: | :: :
Db         523 DLRAIEHSITKLEESLTSLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYVDHSGVIKD 582

Qy         417 KVKEIRDRIQCRAEELQNTERWGLLSQWM---PWVLPFL----GPLAALILLLLFGPCIF 469
                : ::|:|:  |  | :: : |     |    ||:   |    |||  |:||| |||||
Db         583 SMAKLRERLDIRQRERESKQGW--FESWFNKSPWLTTLLSTIAGPLIILLLLLTFGPCIL 640

Qy         470 NLLVKFVSSRIEAVKLQMVLQME 492
               | || |:  || ||:: |||: :
Db         641 NKLVAFIRERINAVQV-MVLKQQ 662

RESULT 14
US-09-315-127-11
; Sequence 11, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
;  APPLICANT: The University of Tennessee, c/o Richard Cox
;  TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
;  TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
;  FILE REFERENCE: 44137-5023, U. of Tennessee
;  CURRENT APPLICATION NUMBER: US/09/315,127
;  CURRENT FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 654
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-315-127-11

  Query Match             14.6%;  Score 385;  DB 4;  Length 654;
  Best Local Similarity   28.4%;  Pred. No. 1.9e-30;
  Matches  108;  Conservative   60;  Mismatches  154;  Indels   58;  Gaps   10;

Qy        141 STPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISI 200
              | | |  |           |  ||::|       |:  :      ||:||      |  :
Db        291 SVPQPPPG-----------TGDRLLALVKGAYQALNLTNPDKTQECWLCLVSGPPYYEGV 339

Qy        201 PVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTR 260
               |   : | ||             :|   |    ||  :      |    |:  |    :
Db        340 AVVGTYTNHST-----------APANCTATSQHKLTLSEV-----TGQGLCMGAVPKTHQ 383

Qy        261 IVC--------------LPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTE 302
               :|               |:|  : | |    ||: :    :   | |  | | : ::
Db        384 ALCNTTQSAGSGSYYLAAPAGTMWACSTGLTPCLSTTVLNLTTDYCVLVELWPRVIYHSP 443

Qy        303 QDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTGIGSITTS---TQFYYKLSQEING 356
                :|  :  :   || |:   |  ::    :| :  |||: ||:   || : :|   |
Db        444 DYMYGQLEQRTKYKREPVSLTLALLLGGLTMGGIAAGIGTGTTALIKTQQFEQLHAAIQT 503

Qy        357 DMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTE 416
              |: :|  |:  |:  | ||: ||||||| ||||  | || |  | || |:| : : :| :
Db        504 DLNEVEKSITNLEKSLTSLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRD 563

Qy        417 KVKEIRDRIQCRAEELQNTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFN 470
               : ::|:|:  | :  :  : |  || ::  ||    :   :|||  |:|:||||||| |
Db        564 SMAKLRERLNQRQKLFETGQGWFEGLFNR-SPWFTTLISTIMGPLIVLLLILLFGPCILN 622

Qy        471 LLVKFVSSRIEAVKLQMVLQ 490
               ||:||  ||  |:  :: |
Db        623 RLVQFVKDRISVVQALVLTQ 642


RESULT 15
US-09-315-127-12
; Sequence 12, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
;  APPLICANT: The University of Tennessee, c/o Richard Cox
;  TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
;  TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
;  FILE REFERENCE: 44137-5023, U. of Tennessee
;  CURRENT APPLICATION NUMBER: US/09/315,127
;  CURRENT FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 654
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SEQ. ID NO.
;   OTHER INFORMATION: 8, envelope protein produced by retroviral vector
;   OTHER INFORMATION: of seq. id no. 7
US-09-315-127-12

  Query Match             14.6%;  Score 385;  DB 4;  Length 654;
  Best Local Similarity   28.4%;  Pred. No. 1.9e-30;
  Matches  108;  Conservative   60;  Mismatches  154;  Indels   58;  Gaps   10;

Qy        141 STPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISI 200
              | | |  |           |  ||::|       |:  :      ||:||      |  :
Db        291 SVPQPPPG-----------TGDRLLALVKGAYQALNLTNPDKTQECWLCLVSGPPYYEGV 339

Qy        201 PVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTR 260
               |   : | ||             :|   |    ||  :      |    |:  |    :
Db        340 AVVGTYTNHST-----------APANCTATSQHKLTLSEV-----TGQGLCMGAVPKTHQ 383

Qy        261 IVC--------------LPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTE 302
               :|               |:|  : | |    ||: :    :   | |  | | : ::
Db        384 ALCNTTQSAGSGSYYLAAPAGTMWACSTGLTPCLSTTVLNLTTDYCVLVELWPRVIYHSP 443

Qy        303 QDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTGIGSITTS---TQFYYKLSQEING 356
                :|  :  :   || |:   |  ::    :| :  |||: ||:   || : :|   |
Db        444 DYMYGQLEQRTKYKREPVSLTLALLLGGLTMGGIAAGIGTGTTALIKTQQFEQLHAAIQT 503

Qy        357 DMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTE 416
              |: :|  |:  |:  | ||: ||||||| ||||  | || |  | || |:| : : :| :
Db        504 DLNEVEKSITNLEKSLTSLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRD 563

Qy        417 KVKEIRDRIQCRAEELQNTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFN 470
               : ::|:|:  | :  :  : |  || ::  ||    :   :|||  |:|:||||||| |
Db        564 SMAKLRERLNQRQKLFETGQGWFEGLFNR-SPWFTTLISTIMGPLIVLLLILLFGPCILN 622

Qy        471 LLVKFVSSRIEAVKLQMVLQ 490
               ||:||  ||  |:  :: |
Db        623 RLVQFVKDRISVVQALVLTQ 642


RESULT 16
US-09-315-127-8
; Sequence 8, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
;  APPLICANT: The University of Tennessee, c/o Richard Cox
;  TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
;  TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
;  FILE REFERENCE: 44137-5023, U. of Tennessee
;  CURRENT APPLICATION NUMBER: US/09/315,127
;  CURRENT FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 645
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
US-09-315-127-8

  Query Match             14.3%;  Score 376.5;  DB 4;  Length 645;
  Best Local Similarity   28.4%;  Pred. No. 1.4e-29;
  Matches  103;  Conservative   63;  Mismatches  146;  Indels   51;  Gaps   10;

Qy        160 THTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFST-------- 211
              |  ||::|       |:  :      ||:||      |  : |   : | ||
Db        290 TGDRLLNLVEGAYQALNLTNPDKTQECWLCLVSGPPYYEGVAVVGTYTNHSTAPASCTAT 349

Qy        212 --------EINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC 263
                      |:    : :| |      || :        ||  :  |         :  :
Db        350 SQHKLTLSEVTGQGLCMGALPK----THQA------LCNTTQSAGS--------GSYYLA 391

Qy        264 LPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVP 319
               |:|  : | |    ||: :    :   | |  | | :  ::   :|  :  :   || |
Db        392 APAGTMWACSTGLTPCLSTTMLNLTTDYCVLVELWPRIIYHSPDYMYGQLEQRTKYKREP 451

Qy        320 I---LPFVIRAGVLGRLGTGIGSITTS---TQFYYKLSQEINGDMEQVTDSLVTLQDQLN 373
              :   |  ::    :| :  |||: ||:   || : :|   |  |: :|  |:  |:  |
Db        452 VSLTLALLLGGLTMGGIAAGIGTGTTALIKTQQFEQLHAAIQTDLNEVEKSITNLEKSLT 511

Qy        374 SLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQ 433
              ||: ||||||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  :
Db        512 SLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFE 571

Qy        434 NTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQM 487
              : : |  |  ::  ||    :   :|||  |:|:||||||| | ||:||  ||  |:  :
Db        572 SGQGWFEGQFNR-SPWFTTLISTIMGPLIVLLLILLFGPCILNRLVQFVKDRISVVQALV 630
```

```
Qy        488 VLQ 490
              : |
Db        631 LTQ 633


RESULT 17
US-09-315-127-9
; Sequence 9, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
;  APPLICANT: The University of Tennessee, c/o Richard Cox
;  TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
;  TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
;  FILE REFERENCE: 44137-5023, U. of Tennessee
;  CURRENT APPLICATION NUMBER: US/09/315,127
;  CURRENT FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 645
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SEQ. ID NO.
;   OTHER INFORMATION: 6, envelope protein produced by retroviral vector
;   OTHER INFORMATION: of seq. id no. 5
US-09-315-127-9

  Query Match             14.3%;  Score 376.5;  DB 4;  Length 645;
  Best Local Similarity   28.4%;  Pred. No. 1.4e-29;
  Matches  103;  Conservative   63;  Mismatches  146;  Indels   51;  Gaps   10;

Qy        160 THTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFST-------- 211
              |  ||::|       |:  :      ||:||      |  : |   : | ||
Db        290 TGDRLLNLVEGAYQALNLTNPDKTQECWLCLVSGPPYYEGVAVVGTYTNHSTAPASCTAT 349

Qy        212 --------EINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC 263
                      |:    : :| |      || :        ||  :  |         :  :
Db        350 SQHKLTLSEVTGQGLCMGALPK----THQA------LCNTTQSAGS--------GSYYLA 391

Qy        264 LPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVP 319
               |:|  : | |    ||: :    :   | |  | | :  ::   :|  :  :   || |
Db        392 APAGTMWACSTGLTPCLSTTMLNLTTDYCVLVELWPRIIYHSPDYMYGQLEQRTKYKREP 451

Qy        320 I---LPFVIRAGVLGRLGTGIGSITTS---TQFYYKLSQEINGDMEQVTDSLVTLQDQLN 373
              :   |  ::    :| :  |||: ||:   || : :|   |  |: :|  |:  |:  |
Db        452 VSLTLALLLGGLTMGGIAAGIGTGTTALIKTQQFEQLHAAIQTDLNEVEKSITNLEKSLT 511

Qy        374 SLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQ 433
              ||: ||||||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  :
Db        512 SLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFE 571

Qy        434 NTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQM 487
              : : |  |  ::  ||    :   :|||  |:|:||||||| | ||:||  ||  |:  :
Db        572 SGQGWFEGQFNR-SPWFTTLISTIMGPLIVLLLILLFGPCILNRLVQFVKDRISVVQALV 630

Qy        488 VLQ 490
              : |
Db        631 LTQ 633


RESULT 18
US-09-111-085-2
; Sequence 2, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
;  APPLICANT: Stoye, Jonathan P
;  APPLICANT: Weiss, Robin A
;  TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
;  TITLE OF INVENTION: specific sequences
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 660
;   TYPE: PRT
;   ORGANISM: Porcine retrovirus
US-09-111-085-2

  Query Match             14.3%;  Score 376.5;  DB 3;  Length 660;
  Best Local Similarity   27.5%;  Pred. No. 1.4e-29;
  Matches  111;  Conservative   68;  Mismatches  153;  Indels   71;  Gaps   13;

Qy        135 QLTRGHSTPSPY---KGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLP 191
              |  |    || |    | | :: : |::|  :| ||       |:  : :  ::||:||
Db        272 QEQRPSPNPSDYNTTSGSVPTEPNITIKTGAKLFSLIQGAFQALNSTTPEATSSCWLCLA 331

Qy        192 LHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCV-----KFSNTIDT 246
                   |  :    ::|                     :|      |      | : |  :
Db        332 SGPPYYEGMARGGKFN---------------------VTKEHRDQCTWGSQNKLTLTEVS 370

Qy        247 TSSQCIRWVTPPTRIVCLPSGIF----------------FVCGTSAYHCLN----GSSES 286
                  ||  | |  : :|  :  |                : | |    |::      ::
Db        371 GKGTCIGMVPPSHQHLCNHTEAFNRTSESQYLVPGYDRWWACNTGLTPCVSTLVFNQTKD 430

Qy        287 MCFLSFLVPPMTIYTEQ---DLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTGIGSI 340
               | :  :|| :  | |:   | |::   :|  || ||   |  ::  ||   :|||  ::
Db        431 FCVMVQIVPRVYYYPEKAVLDEYDYRYNRP--KREPISLTLAVMLGLGVAAGVGTGTAAL 488

Qy        341 TTSTQFYYK----LSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGT 396
               |  |   |    | : :  |:: :  |:  |:: | ||: ||||||| ||||  | ||
Db        489 ITGPQQLEKGLSNLHRIVTEDLQALEKSVSNLEESLTSLSEVVLQNRRGLDLLFLKEGGL 548

Qy        397 CLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWM---PWVLPFL- 452
              |: | || |:||: |  : : : ::|:|:: | |  : :  |     |    ||:   |
Db        549 CVALKEECCFYVDHSGAIRDSMSKLRERLERRRREREADQGW--FEGWFNRSPWMTTLLS 606

Qy        453 ---GPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQME 492
                 |||  |:|||  |||: |  |  ||  |: ||:: |||: :
Db        607 ALTGPLVVLLLLLTVGPCLINRFVAFVRERVSAVQI-MVLRQQ 648


RESULT 19
US-09-376-781-5
; Sequence 5, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 660
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:PERV-A
;   OTHER INFORMATION: polypeptide sequence taken from GenBank Accession
;   OTHER INFORMATION: No. 6261806 Y12238 for comparison.
US-09-376-781-5

  Query Match             14.3%;  Score 376.5;  DB 3;  Length 660;
  Best Local Similarity   27.5%;  Pred. No. 1.4e-29;
  Matches  111;  Conservative   68;  Mismatches  153;  Indels   71;  Gaps   13;

Qy          135 QLTRGHSTPSPY---KGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLP 191
                |  |    || |    | | :: : |::|  :| ||       |:  : :  ::||:||
Db          272 QEQRPSPNPSDYNTTSGSVPTEPNITIKTGAKLFSLIQGAFQALNSTTPEATSSCWLCLA 331

Qy          192 LHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCV-----KFSNTIDT 246
                     |  :    ::|                     :|      |      | : |  :
Db          332 SGPPYYEGMARGGKFN---------------------VTKEHRDQCTWGSQNKLTLTEVS 370

Qy          247 TSSQCIRWVTPPTRIVCLPSGIF----------------FVCGTSAYHCLN----GSSES 286
                    ||  | |  : :|  :  |                : | |    |::      ::
Db          371 GKGTCIGMVPPSHQHLCNHTEAFNRTSESQYLVPGYDRWWACNTGLTPCVSTLVFNQTKD 430

Qy          287 MCFLSFLVPPMTIYTEQ---DLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTGIGSI 340
                 | :  :|| :  | |:   | |::   :|  || ||   |  ::  ||   :|||  ::
Db          431 FCVMVQIVPRVYYYPEKAVLDEYDYRYNRP--KREPISLTLAVMLGLGVAAGVGTGTAAL 488

Qy          341 TTSTQFYYK----LSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGT 396
                 |  |   |    | : :  |:: :  |:  |:: | ||: ||||||| ||||  | ||
Db          489 ITGPQQLEKGLSNLHRIVTEDLQALEKSVSNLEESLTSLSEVVLQNRRGLDLLFLKEGGL 548

Qy          397 CLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWM---PWVLPFL- 452
                |: || |:||: |  : : : ::|:|:: |  | :  : |     |    ||:   |
Db          549 CVALKEECCFYVDHSGAIRDSMSKLRERLERRRREREADQGW--FEGWFNRSPWMTTLLS 606

Qy          453 ---GPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQME 492
                   |||  |:|||  |||: |  | ||  |: ||:: |||: :
Db          607 ALTGPLVVLLLLLTVGPCLINRFVAFVRERVSAVQI-MVLRQQ 648

RESULT 20
US-09-376-781-6
; Sequence 6, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 638
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:PERV-C
;   OTHER INFORMATION: polypeptide sequence taken from GenBank Accession
;   OTHER INFORMATION: No. 6261806 AF038600 for comparison.
US-09-376-781-6

  Query Match             14.0%;  Score 368;  DB 3;  Length 638;
  Best Local Similarity   24.8%;  Pred. No. 1e-28;
  Matches  132;  Conservative   82;  Mismatches  187;  Indels  132;  Gaps   21;

Qy           65 HTHMPRNCYNSATLCMHANTHYW----------TGKMINPSCPGGLGATVCW--TYFTHT 112
                |   ||: :     |: :|  ||          :   :|     |    : |  |
Db          122 HCGNPRDFFCKQWNCVTSNDGYWKWPTSQQDRVSFSYVNTYTSSGQFNYLTWIRTGSPKC 181

Qy          113 SMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKG-------------LVLSKLH---- 155
                | ||   ::    ||  :| | :   | |    | |             | :::|
Db          182 SPSDLDYLKISFTEKGKQENILKWVNGMSWGMVYYGGSGKQPGSILTIRLKINQLEPPMA 241

Qy          156 ---------------------------------ETLRTHTRLVSLFNTTLTRLHEVSAQN 182
                                                  | :   :| ||       |:  : :
Db          242 IGPNTVLTGQRPPTQGPGPSSNITSGSDPTESSSTTKMGAKLFSLIQGAFQALNSTTPEA 301

Qy          183 PTNCWMCL---PLHF-----RPYISIPVPEQ----WNNFS----TEINTTSVLVGPLVSN 226
                 ::||:||   | ::     |   ::    :    | : :    ||::     :| :  :
Db          302 TSSCWLCLASGPPYYEGMARRGKFNVTKEHRDQCTWGSQNKLTLTEVSGKGTCIGKVPPS 361

Qy          227 LEITHTSNLTCVKFSNTIDTTSSQCI-----RWVTPPTRIVCLPSGIFFVCGTSAYHCLN 281
                 :  |  | |   |:    |: || :     ||               : | |    |::
Db          362 HQ--HLCNHT-EAFN---QTSESQYLVPGYDRW---------------WACNTGLTPCVS 400

Qy          282 ----GSSESMCFLSFLVPPMTIYTEQ------DLYNHVVPKPHNKRVPI---LPFVIRAG 328
                      ::  | :  :|| :  | |:      |  ||       || ||   |  ::  |
Db          401 TLVFNQTKDFCIMVQIVPRVYYYPEKAILDEYDYRNH-----RQKREPISLTLAVMLGLG 455

Qy          329 VLGRLGTGIGSITTSTQ----FYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRR 384
                |   :|||  :: |  |        | : :  |:: :  |:  |:: | ||: |||||||
Db          456 VAAGVGTGTAALVTGPQQLETGLSNLHRIVTEDLQALEKSVSNLEESLTSLSEVVLQNRR 515

Qy          385 ALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNT----ERWGL 440
                 ||||  | || |: | || |:||: |  : : : ::|:|:: |  | : |    | |
Db          516 GLDLLFLKEGGLCVALKEECCFYVDHSGAIRDSMNKLRERLEKRRREKETTQGWFEGWFN 575

Qy          441 LSQWMPWVLPFL-GPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQME 492
                 | |:  :|  | |||  |:|||  |||| | |: |:  || ||:: |||: :
Db          576 RSLWLATLLSALTGPLIVLLLLLTVGPCIINKLIAFIRERISAVQI-MVLRQQ 627

RESULT 21
US-08-486-099-107
; Sequence 107, Application US/08486099
; Patent No. 6013263
;  GENERAL INFORMATION:
;    APPLICANT:  Bolognesi, Dani P.
;    APPLICANT:  Matthews, Thomas J.
;    APPLICANT:  Wild, Carl T.
;    APPLICANT:  Barney, Shawn O.
;    APPLICANT:  Lambert, Dennis M.
;    APPLICANT:  Petteway, Stephen R.
;    APPLICANT:  Langlois, Alphonse J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION OF
;    TITLE OF INVENTION:  MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS
;    TITLE OF INVENTION:  B VIRUS TRANSMISSION
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,099
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7872-031
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  107:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  192 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-486-099-107

  Query Match            13.8%;  Score 364;  DB 3;  Length 192;
  Best Local Similarity  39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
              : ::|  :  |:   : ||   :  |   | ||| ::  |:: :: ::  ||||::|||
Db          2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy        378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
              ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db         62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121

Qy        438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
              |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db        122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 22
US-08-360-107A-117
; Sequence 117, Application US/08360107A
; Patent No. 6017536
;  GENERAL INFORMATION:
;    APPLICANT:  Bolognesi, Dani P.
;    APPLICANT:  Matthews, Thomas J.
;    APPLICANT:  Wild, Carl T.
;    APPLICANT:  Barney, Shawn O.
;    APPLICANT:  Lambert, Dennis M.
;    APPLICANT:  Petteway, Stephen R.
;    APPLICANT:  Langlois, Alphonse J.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR INHIBITION
;    TITLE OF INVENTION:  OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
;    TITLE OF INVENTION:  TRANSMISSION
;    NUMBER OF SEQUENCES:  149
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/360,107A
;      FILING DATE:  20-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7872-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  117:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  192 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-360-107A-117

  Query Match            13.8%;  Score 364;  DB 3;  Length 192;
  Best Local Similarity  39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
              : ::|  :  |:   : ||   :  |   | ||| ::  |:: :: ::  ||||::|||
Db          2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy        378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
              ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db         62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121

Qy        438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
              |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db        122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 23
US-08-484-223B-107
; Sequence 107, Application US/08484223B
; Patent No. 6020459
;  GENERAL INFORMATION:
;    APPLICANT:  Bolognesi, Dani P.
;    APPLICANT:  Matthews, Thomas J.
;    APPLICANT:  Wild, Carl T.
;    APPLICANT:  Barney, Shawn O.
;    APPLICANT:  Lambert, Dennis M.
;    APPLICANT:  Petteway, Stephen R.
;    APPLICANT:  Langlois, Alphonse J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION OF
;    TITLE OF INVENTION:  MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
;    TITLE OF INVENTION:  TRANSMISSION
;    NUMBER OF SEQUENCES:  245
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,223B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7872-029
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  107:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  192 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-484-223B-107

  Query Match            13.8%;  Score 364;  DB 3;  Length 192;
  Best Local Similarity  39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
              : ::|  :  |:   : ||   :  |   | ||| ::  |:: :: ::  ||||::|||
Db          2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy        378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
              ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db         62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121
```

```
Qy          438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
                |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db          122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172


RESULT 24
US-08-919-597-107
; Sequence 107, Application US/08919597
; Patent No. 6054265
;  GENERAL INFORMATION:
;    APPLICANT:  Bolognesi, Dani P.
;    APPLICANT:  Matthews, Thomas J.
;    APPLICANT:  Wild, Carl T.
;    APPLICANT:  Barney, Shawn O.
;    APPLICANT:  Lambert, Dennis M.
;    APPLICANT:  Petteway, Stephen R.
;    APPLICANT:  Langlois, Alphonse J.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR INHIBITION
;    TITLE OF INVENTION:  OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
;    TITLE OF INVENTION:  TRANSMISSION
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/919,597
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/470,896
;      FILING DATE:  06-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7872-020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  107:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  192 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-919-597-107

  Query Match             13.8%;  Score 364;  DB 3;  Length 192;
  Best Local Similarity   39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
                : ::|  :  |:   : ||   :  |   | |||::   |:: :: ::  ||||::|||
Db            2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy          378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
                ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db           62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121

Qy          438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
                |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db          122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172


RESULT 25
US-08-475-668A-107
; Sequence 107, Application US/08475668A
; Patent No. 6060065
;  GENERAL INFORMATION:
;    APPLICANT:  Barney, Shawn O.
;    APPLICANT:  Lambert, Dennis M.
;    APPLICANT:  Petteway, Stephen R.
;    TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION OF MEMBRANE
;    TITLE OF INVENTION:  FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
;    TITLE OF INVENTION:  TRANSMISSION
;    NUMBER OF SEQUENCES:  211
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/475,668A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7872-026
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  107:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  192 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-475-668A-107

  Query Match             13.8%;  Score 364;  DB 3;  Length 192;
  Best Local Similarity   39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
                : ::|  :  |:   : ||   :  |   | |||::   |:: :: ::  ||||::|||
Db            2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy          378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
                ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db           62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121

Qy          438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
                |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db          122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172


RESULT 26
US-08-485-551A-107
; Sequence 107, Application US/08485551A
; Patent No. 6068973
;  GENERAL INFORMATION:
;    APPLICANT:  Bolognesi, Dani P.
;    APPLICANT:  Matthews, Thomas J.
;    APPLICANT:  Wild, Carl T.
;    APPLICANT:  Barney, Shawn O.
;     APPLICANT:  Lambert, Dennis M.
;     APPLICANT:  Petteway, Stephen R.
;     APPLICANT:  Langlois, Alphonse J.
;     TITLE OF INVENTION:  METHODS FOR INHIBITION OF MEMBRANE
;     TITLE OF INVENTION:  FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
;     TITLE OF INVENTION:  TRANSMISSION
;     NUMBER OF SEQUENCES:  211
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/485,551A
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7872-023
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  107:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  192 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-08-485-551A-107

  Query Match             13.8%;  Score 364;  DB 3;  Length 192;
  Best Local Similarity   39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
              : ::|  :  |:   : ||   :  |   | ||| ::  |:: :: ::  ||||::|||
Db          2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy        378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
              ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db         62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121

Qy        438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
              |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db        122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172


RESULT 27
US-08-471-913A-107
; Sequence 107, Application US/08471913A
; Patent No. 6093794
;   GENERAL INFORMATION:
;     APPLICANT:  Bolognesi, Dani P.
;     APPLICANT:  Matthews, Thomas J.
;     APPLICANT:  Wild, Carl T.
;     APPLICANT:  Barney, Shawn O.
;     APPLICANT:  Lambert, Dennis M.
;     APPLICANT:  Petteway, Stephen R.
;     APPLICANT:  Langlois, Alphonse J.
;     TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION OF MEMBRANE
;     TITLE OF INVENTION:  FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
;     TITLE OF INVENTION:  TRANSMISSION
;     NUMBER OF SEQUENCES:  214
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10036-2711
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/471,913A
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7872-030
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  107:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  192 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
US-08-471-913A-107

  Query Match             13.8%;  Score 364;  DB 3;  Length 192;
  Best Local Similarity   39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
              : ::|  :  |:   : ||   :  |   | ||| ::  |:: :: ::  ||||::|||
Db          2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy        378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
              ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db         62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121

Qy        438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
              |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db        122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172


RESULT 28
US-08-485-264A-107
; Sequence 107, Application US/08485264A
; Patent No. 6228983
;   GENERAL INFORMATION:
;     APPLICANT:  Bolognesi, Dani P.
;     APPLICANT:  Matthews, Thomas J.
;     APPLICANT:  Wild, Carl T.
;     APPLICANT:  Barney, Shawn O.
;     APPLICANT:  Lambert, Dennis M.
;     APPLICANT:  Petteway, Stephen R.
;     APPLICANT:  Langlois, Alphonse J.
;     TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION OF
;     TITLE OF INVENTION:  MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
;     TITLE OF INVENTION:  RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
;     NUMBER OF SEQUENCES:  232
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds LLP
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,264A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7872-021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  107:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  192 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-485-264A-107

  Query Match             13.8%;  Score 364;  DB 3;  Length 192;
  Best Local Similarity   39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
              : ::|  :  |:   : ||   :  |   | ||| ::  |:: :: ::  ||||::|||
Db          2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy        378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
              ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db         62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121

Qy        438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
              |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db        122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 29
US-08-474-349A-107
; Sequence 107, Application US/08474349A
; Patent No. 6333395
;  GENERAL INFORMATION:
;    APPLICANT:  Bolognesi, Dani P.
;    APPLICANT:  Matthews, Thomas J.
;    APPLICANT:  Wild, Carl T.
;    APPLICANT:  Barney, Shawn O.
;    APPLICANT:  Lambert, Dennis M.
;    APPLICANT:  Petteway, Stephen R.
;    APPLICANT:  Langlois, Alphonse J.
;    TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION OF MEMBRANE
;    TITLE OF INVENTION:  FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
;    TITLE OF INVENTION:  VIRUS TRANSMISSION
;    NUMBER OF SEQUENCES:  517
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,349A
```

```
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7872-024
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  107:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  192 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-474-349A-107

  Query Match             13.8%;  Score 364;  DB 3;  Length 192;
  Best Local Similarity   39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
              : ::|  :  |:   : ||   :  |   | ||| ::  |:: :: ::  ||||::|||
Db          2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy        378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
              ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db         62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121

Qy        438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
              |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db        122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 30
US-08-470-896-107
; Sequence 107, Application US/08470896
; Patent No. 6479055
;  GENERAL INFORMATION:
;    APPLICANT:  Bolognesi, Dani P.
;    APPLICANT:  Matthews, Thomas J.
;    APPLICANT:  Wild, Carl T.
;    APPLICANT:  Barney, Shawn O.
;    APPLICANT:  Lambert, Dennis M.
;    APPLICANT:  Petteway, Stephen R.
;    APPLICANT:  Langlois, Alphonse J.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR INHIBITION
;    TITLE OF INVENTION:  OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
;    TITLE OF INVENTION:  TRANSMISSION
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,896
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7872-020
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  107:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  192 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-470-896-107

  Query Match             13.8%;  Score 364;  DB 4;  Length 192;
  Best Local Similarity   39.8%;  Pred. No. 3.9e-29;
  Matches   68;  Conservative   44;  Mismatches   59;  Indels    0;  Gaps    0;

Qy        318 VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
              : ::|  :  |:   : ||   :  |   | ||| ::  |:: :: ::  ||||::|||
Db          2 IQLIPLFVGLGITTAVSTGAAGLGVSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

Qy        378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
              ||||||| ||||||::|| || | |: |:| |:| || :|:| ::| :: |  :| :
Db         62 VVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPF 121

Qy        438 WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
              |     ::|:|:| ||||  |:|:| ||| ||| |: |:  :||::: : :
Db        122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172


Search completed: January 22, 2005, 22:12:41
Job time : 48 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM protein - protein search, using sw model

Run on:        January 22, 2005, 22:07:21 ; Search time 106 Seconds
                                            (without alignments)
                                            1680.336 Million cell updates/sec

Title:         US-09-319-156B-10
Perfect score: 2634
Sequence:      1 MALPYHTFLFTVLLPPFALT..........KFVSSRIEAVKLQMVLQMEP 493

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters:      1608061

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
               1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
               2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
               3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
               4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
               5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
               6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
               7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
               8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
               9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
               10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
               11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
               12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
               13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
               14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
               15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
               16:  /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
               17:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
               18:  /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
               19:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
               20:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                    %
Result              Query
  No.    Score     Match Length DB  ID                            Description
 -------------------------------------------------------------------------------
     1    2632      99.9    493  8  US-08-979-847-106            Sequence 106, App
     2    2632      99.9    493 14  US-10-114-104-106            Sequence 106, App
     3    2447      92.9    542 15  US-10-416-642-2              Sequence 2, Appli
     4    2419      91.8    503 15  US-10-637-565-20             Sequence 20, Appl
     5    2319      88.0    538  9  US-09-902-535-2              Sequence 2, Appli
     6    2319      88.0    538 13  US-10-114-893-135            Sequence 135, App
     7    2319      88.0    538 14  US-10-016-249-4              Sequence 4, Appli
     8    2314      87.9    538 15  US-10-133-036-2              Sequence 2, Appli
     9    2285      86.8    538 15  US-10-133-036-21             Sequence 21, Appl
    10    2237      84.9    538 15  US-10-133-036-20             Sequence 20, Appl
    11  1235.5      46.9    263  9  US-09-864-761-37656          Sequence 37656, A
    12    1122      42.6    531 15  US-10-363-616-473            Sequence 473, App
    13     591      22.4    120 15  US-10-133-036-4              Sequence 4, Appli
    14     516      19.6    162  8  US-08-979-847-109            Sequence 109, App
    15     516      19.6    162 14  US-10-114-104-109            Sequence 109, App
    16   514.5      19.5    576 10  US-09-793-451-4              Sequence 4, Appli
    17   514.5      19.5    576 14  US-10-283-722-4              Sequence 4, Appli
    18   514.5      19.5    576 14  US-10-283-903-4              Sequence 4, Appli
    19   480.5      18.2    469 13  US-10-001-835-168            Sequence 168, App
    20     452      17.2    433  8  US-08-979-847-81             Sequence 81, Appl
    21     452      17.2    433 14  US-10-430-442-87             Sequence 87, Appl
    22     452      17.2    433 14  US-10-114-104-81             Sequence 81, Appl
    23     423      16.1    253 10  US-09-764-891-2970           Sequence 2970, Ap
    24     421      16.0    618 15  US-10-276-774-1749           Sequence 1749, Ap
    25     393      14.9    676  9  US-09-954-983-11             Sequence 11, Appl
    26     392      14.9    675  9  US-09-954-983-8              Sequence 8, Appli
    27     391      14.8    672  9  US-09-954-983-4              Sequence 4, Appli
    28     389      14.8    675  9  US-09-954-983-6              Sequence 6, Appli
    29     386      14.7    672  9  US-09-954-983-5              Sequence 5, Appli
    30     386      14.7    675  9  US-09-954-983-7              Sequence 7, Appli
    31     378      14.4    638 16  US-10-441-949-32             Sequence 32, Appl
    32     378      14.4    638 16  US-10-441-949-34             Sequence 34, Appl
    33     378      14.4    638 16  US-10-441-949-40             Sequence 40, Appl
    34     378      14.4    638 16  US-10-441-949-42             Sequence 42, Appl
    35     378      14.4    638 16  US-10-441-949-45             Sequence 45, Appl
    36     378      14.4    678 16  US-10-441-949-36             Sequence 36, Appl
    37     378      14.4    678 16  US-10-441-949-38             Sequence 38, Appl
    38   376.5      14.3    660  9  US-09-851-859A-5             Sequence 5, Appli
    39   376.5      14.3    660 16  US-10-441-949-8              Sequence 8, Appli
    40   376.5      14.3    660 16  US-10-441-949-10             Sequence 10, Appl
    41   376.5      14.3    660 16  US-10-441-949-16             Sequence 16, Appl
    42   376.5      14.3    660 16  US-10-441-949-18             Sequence 18, Appl
    43   376.5      14.3    661 16  US-10-441-949-43             Sequence 43, Appl
    44   374.5      14.2    653 14  US-10-029-656-4              Sequence 4, Appli
    45   372.5      14.1    653 14  US-10-029-656-2              Sequence 2, Appli


                                  ALIGNMENTS


RESULT 1
US-08-979-847-106
; Sequence 106, Application US/08979847
; Publication No. US20030039664A1
;  GENERAL INFORMATION:
;    APPLICANT:  PERRON, HERVE
;    APPLICANT:  BESEME, FREDERIC
;    APPLICANT:  BEDIN, FREDERIC
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  KOMURIAN-PRADEL, FLORENCE
;    APPLICANT:  JOLIVET-REYNAUD, COLETTE
;    APPLICANT:  MANDRAND, BERNARD
;    APPLICANT:  GARSON, JEREMY
;    APPLICANT:  TUKE, PHILIP
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  210
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/979,847
;      FILING DATE:  26-NOV-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERRIDGE, WILLIAM P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 39046A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  106:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  493 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-979-847-106

  Query Match             99.9%;  Score 2632;  DB 8;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-240;
  Matches  493;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              |||||||||||||
Db        481 EAVKLQMVLQMEP 493


RESULT 2
US-10-114-104-106
; Sequence 106, Application US/10114104
; Publication No. US20030198647A1
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/114,104
;            FILING DATE: 03-Apr-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/979,847
;            FILING DATE: 26-NOV-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;  INFORMATION FOR SEQ ID NO: 106:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 493 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-10-114-104-106

  Query Match             99.9%;  Score 2632;  DB 14;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-240;
  Matches  493;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
```

```
              |||||||||||||
Db        481 EAVKLQMVLQMEP 493


RESULT 3
US-10-416-642-2
; Sequence 2, Application US/10416642
; Publication No. US20040043452A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  RAMKUMAR, Jayalaxmi
;  APPLICANT:  ARVIZU, Chandra
;  TITLE OF INVENTION: EMBRYOGENESIS ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-0842 PCT
;  CURRENT APPLICATION NUMBER: US/10/416,642
;  CURRENT FILING DATE:  2003-05-13
;  PRIOR APPLICATION NUMBER: 60/249,407
;  PRIOR FILING DATE: 2000-11-15
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PERL Program
; SEQ ID NO 2
;   LENGTH: 542
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20040043452A1 7477736CD1
US-10-416-642-2

  Query Match             92.9%;  Score 2447;  DB 15;  Length 542;
  Best Local Similarity   93.1%;  Pred. No. 2.1e-222;
  Matches  459;  Conservative    8;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||  |||||| |||:|||| || | ||||||| || | | | :|||||||||||||
Db          1 MALPYCIFLFTVLSPPFSLTAPSPCHCRTSSSPYQAFLWRMRRPRHIDAPSYRSLSKGNP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
               |||||||| |||||||||||||||||||||||||||||||||:|||||||| ||||||:
Db         61 AFTAHTHMPHNCYNSATLCMHANTHYWTGKMINPSCPGGLGATICWTYFTHTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| |||||||||| |||||||||||| |||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHSTPSPYKGLDLSKLHETLRTHTHLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||||||||||||||:|||||||||||||||||:||||||||||| ||||||||||
Db        181 QNPTNCWMCLPLHFRPYISVPVPEQWNNFSTEINTTSILVGPLVSNLEIIHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||| ||:|||||||||||:||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIATTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEECCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        421 IRDRIQCRAEELQNTEHWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              |||||||||||||
Db        481 EAVKLQMVLQMEP 493


RESULT 4
US-10-637-565-20
; Sequence 20, Application US/10637565
```

```
; Publication No. US20040043381A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  PERRON, Herve
;  APPLICANT:  KOMURIAN-PRADEL, Florence
;  TITLE OF INVENTION: THE LTR REGION OF MSRV-1 AND THE PROTEINS IT ENCODES, AND PROBES
;  TITLE OF INVENTION:  METHODS FOR DETECTING MSRV-1 RETROVIRUS
;  FILE REFERENCE: 110257
;  CURRENT APPLICATION NUMBER: US/10/637,565
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: US/09/890,340
;  PRIOR FILING DATE: 2001-11-27
;  PRIOR APPLICATION NUMBER: PCT/IB00/00159
;  PRIOR FILING DATE: 2000-02-15
;  PRIOR APPLICATION NUMBER: EP 99420041.8
;  PRIOR FILING DATE: 1999-02-15
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
;   LENGTH: 503
;   TYPE: PRT
;   ORGANISM: MSRV-1 retrovirus
US-10-637-565-20

  Query Match             91.8%;  Score 2419;  DB 15;  Length 503;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-220;
  Matches  454;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         40 RTRLPGNIDAPSYRSLSKGNSTFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGG 99
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RTRLPGNIDAPSYRSLSKGNSTFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGG 60

Qy        100 LGATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLR 159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LGATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLR 120

Qy        160 THTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVL 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 THTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVL 180

Qy        220 VGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHC 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHC 240

Qy        280 LNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGS 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGS 300

Qy        340 ITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLF 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLF 360

Qy        400 LGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALI 459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALI 420

Qy        460 LLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
              ||||||||||||||||||||||||||||||||||
Db        421 LLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 454


RESULT 5
US-09-902-535-2
; Sequence 2, Application US/09902535
; Patent No. US20020102530A1
; GENERAL INFORMATION:
;  APPLICANT: Keith, Jr., James C.
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  Mi, Sha
;  TITLE OF INVENTION: Methods and compositions for diagnosing
;  TITLE OF INVENTION:  and treating preeclampsia and gestational trophoblast
;  TITLE OF INVENTION:  disorders
;  FILE REFERENCE: GIN-6006B4
;  CURRENT APPLICATION NUMBER: US/09/902,535
;  CURRENT FILING DATE:  2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,657
;  PRIOR FILING DATE: 2000-07-06
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-902-535-2

  Query Match             88.0%;  Score 2319;  DB 9;  Length 538;
  Best Local Similarity   88.6%;  Pred. No. 2.7e-210;
  Matches  436;  Conservative   18;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492

RESULT 6
US-10-114-893-135
; Sequence 135, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT: McCoy, John M.
;  APPLICANT: LaVallie, Edward R.
;  APPLICANT: Collins-Racie, Lisa A.
;  APPLICANT: Evans, Cheryl
;  APPLICANT: Merberg, David
;  APPLICANT: Treacy, Maurice
;  APPLICANT: Bowman, Michael R.
;  APPLICANT: Spaulding, Vikki
;  APPLICANT: Carlin-Duckett, McKeough
;  APPLICANT: Kelleher, Kerry S.
;  APPLICANT: Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: GI 6000-10A
;  CURRENT APPLICATION NUMBER: US/10/114,893
```

```
;  CURRENT FILING DATE: 2002-04-02
;  EARLIER APPLICATION NUMBER: 09/413,232
;  EARLIER FILING DATE: 1999-10-06
;  NUMBER OF SEQ ID NOS: 321
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 135
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-114-893-135

  Query Match             88.0%;  Score 2319;  DB 13;  Length 538;
  Best Local Similarity   88.6%;  Pred. No. 2.7e-210;
  Matches  436;  Conservative   18;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492

RESULT 7
US-10-016-249-4
; Sequence 4, Application US/10016249
; Publication No. US20030100053A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Mi, Sha
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6006B.AJ172A
;  CURRENT APPLICATION NUMBER: US/10/016,249
;  CURRENT FILING DATE:  2001-10-30
;  PRIOR APPLICATION NUMBER: US/09/175,928
;  PRIOR FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-016-249-4

  Query Match            88.0%;  Score 2319;  DB 14;  Length 538;
  Best Local Similarity  88.6%;  Pred. No. 2.7e-210;
  Matches  436;  Conservative   18;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492

RESULT 8
US-10-133-036-2
; Sequence 2, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
;  APPLICANT: Conrad, Bernard
;  APPLICANT:  Mach, Bernard
;  TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
;  FILE REFERENCE: 23135-507
;  CURRENT APPLICATION NUMBER: US/10/133,036
;  CURRENT FILING DATE:  2002-04-26
;  PRIOR APPLICATION NUMBER: PCT/EP00/10659
;  PRIOR FILING DATE: 2000-10-30
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Human endogenous retrovirus
US-10-133-036-2

  Query Match            87.9%;  Score 2314;  DB 15;  Length 538;
  Best Local Similarity  88.4%;  Pred. No. 8.1e-210;
  Matches  435;  Conservative   19;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492

RESULT 9
US-10-133-036-21
; Sequence 21, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
;  APPLICANT: Conrad, Bernard
;  APPLICANT:  Mach, Bernard
;  TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
;  FILE REFERENCE: 23135-507
;  CURRENT APPLICATION NUMBER: US/10/133,036
;  CURRENT FILING DATE:  2002-04-26
;  PRIOR APPLICATION NUMBER: PCT/EP00/10659
;  PRIOR FILING DATE: 2000-10-30
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Retrovirus
US-10-133-036-21

  Query Match            86.8%;  Score 2285;  DB 15;  Length 538;
  Best Local Similarity  87.2%;  Pred. No. 4.5e-207;
  Matches  429;  Conservative   21;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||: | | ||||||| | |||||||||| | : |||||||||||| ||
Db          1 MALPYHIFLFTVVSPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLCKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180
```

```
Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||| ||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMAIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | ||||||||||||||||||:|||||| |||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLRNRRALDSLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492

RESULT 10
US-10-133-036-20
; Sequence 20, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
;  APPLICANT: Conrad, Bernard
;  APPLICANT:  Mach, Bernard
;  TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
;  FILE REFERENCE: 23135-507
;  CURRENT APPLICATION NUMBER: US/10/133,036
;  CURRENT FILING DATE:  2002-04-26
;  PRIOR APPLICATION NUMBER: PCT/EP00/10659
;  PRIOR FILING DATE: 2000-10-30
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Retrovirus
US-10-133-036-20

  Query Match             84.9%;  Score 2237;  DB 15;  Length 538;
  Best Local Similarity   85.6%;  Pred. No. 1.6e-202;
  Matches  421;  Conservative   26;  Mismatches   45;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              | |||| || :|| | | ||||||| | |||||: ||| | : |||||||||||||||
Db          1 MGLPYHIFLCSVLSPCFTLTAPPPCRCMTSSSPHPEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||  | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTGVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||:|||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNVEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
```

```
Db        301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | ||||||||||||||||||:||||||||||:||||||||||| ||||||| ||||||:|
Db        361 VADSLVTLQDQLNSLAAVVLRNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVEE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              | ||||  ||||:||  |||||:||||:|||||||||:|||||||||||:||| ||||||
Db        421 IPDRIQRIAEELRNTGPWGLLSRWMPWILPFLGPLAAIILLLLFGPCIFDLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492

RESULT 11
US-09-864-761-37656
; Sequence 37656, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37656
;   LENGTH: 263
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC002346.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =   6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
;   OTHER INFORMATION: EST_HUMAN HIT: AU138405.1, EVALUE 7.00e-90
;   OTHER INFORMATION: SWISSPROT HIT: P10269, EVALUE 1.00e-16
US-09-864-761-37656

  Query Match             46.9%;  Score 1235.5;  DB 9;  Length 263;
  Best Local Similarity   89.4%;  Pred. No. 4e-108;
  Matches  236;  Conservative    8;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
              || :|| |||||  |||| ||||| |||||||| ||| |||||||||| |||||||||||
Db          1 EKHIKEVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60

Qy        186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              ||||||||||||||||:||||||||||||||||||||| |||||||||||||||||||||
Db         61 CWMCLPLHFRPYISIPIPEQWNNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTID 119

Qy        246 TTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDL 305
              | :||||||||||||||||||||||||||||||||||||||:||||||| || |||||||
Db        120 TANSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDL 179

Qy        306 YNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
              ||||:||| ||||||||||| ||||| : |||| ||||||||||||||:||||| | |:|
Db        180 YNHVIPKPRNKRVPILPFVIGAGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTL 239

Qy        366 VTLQDQLNSLAAVVLQNRRALDLL 389
              ||||||||||||||||||||||||
Db        240 VTLQDQLNSLAAVVLQNRRALDLL 263


RESULT 12
US-10-363-616-473
; Sequence 473, Application US/10363616
; Publication No. US20040044181A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 21272-113 (793)
;  CURRENT APPLICATION NUMBER: US/10/363,616
;  CURRENT FILING DATE:  2003-03-03
;  PRIOR APPLICATION NUMBER: 09/654,935
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 490
; SEQ ID NO 473
;   LENGTH: 531
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-363-616-473

  Query Match             42.6%;  Score 1122;  DB 15;  Length 531;
  Best Local Similarity   51.6%;  Pred. No. 6.2e-97;
  Matches  246;  Conservative    6;  Mismatches   29;  Indels  196;  Gaps    5;

Qy         36 EFLXRTRLPGNIDAPSYRSLSKGNSTFTAHTHMPRN------C-------YNSATLCMH- 81
              ||    :   |: | |  | :         |  ||:      |       :   || :|
Db        112 EFGKNKKQRSNLQATSLISCAT-----VLETRSPRSRCWQVQCLVKAILPFQDDTLLLHP 166

Qy         82 -----ANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQL 136
                   |   |  |||||||||||||||||||||||||||||||||
Db        167 PEGKNAAPFYGGGKMINPSCPGGLGATVCWTYFTHTSMSDGGGIQ--------------- 211

Qy        137 TRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRP 196

Db        212 ------------------------------------------------------------ 211
```

```
Qy        197 YISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVT 256

Db        212 ------------------------------------------------------------ 211

Qy        257 PPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNK 316
                                                   |||||||||||||||||||||||
Db        212 -------------------------------------VPPMTIYTEQDLYNHVVPKPHNK 234

Qy        317 RVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLA 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 RVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLA 294

Qy        377 AVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTE 436
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        295 AVVLQNRRALDLLTAKRGGTCLFLGEECCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTE 354

Qy        437 RWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 HWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 411


RESULT 13
US-10-133-036-4
; Sequence 4, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
;  APPLICANT: Conrad, Bernard
;  APPLICANT:  Mach, Bernard
;  TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
;  FILE REFERENCE: 23135-507
;  CURRENT APPLICATION NUMBER: US/10/133,036
;  CURRENT FILING DATE:  2002-04-26
;  PRIOR APPLICATION NUMBER: PCT/EP00/10659
;  PRIOR FILING DATE: 2000-10-30
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 120
;   TYPE: PRT
;   ORGANISM: Human endogenous retrovirus
US-10-133-036-4

  Query Match             22.4%;  Score 591;  DB 15;  Length 120;
  Best Local Similarity   86.7%;  Pred. No. 1.4e-47;
  Matches  104;  Conservative    3;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||  
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120


RESULT 14
US-08-979-847-109
; Sequence 109, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
;    APPLICANT:  PERRON, HERVE
;    APPLICANT:  BESEME, FREDERIC
;    APPLICANT:  BEDIN, FREDERIC
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  KOMURIAN-PRADEL, FLORENCE
;    APPLICANT:  JOLIVET-REYNAUD, COLETTE
;    APPLICANT:  MANDRAND, BERNARD
;    APPLICANT:  GARSON, JEREMY
;    APPLICANT:  TUKE, PHILIP
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  210
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/979,847
;      FILING DATE:  26-NOV-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERRIDGE, WILLIAM P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 39046A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  109:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  162 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-979-847-109

  Query Match            19.6%;  Score 516;  DB 8;  Length 162;
  Best Local Similarity  87.6%;  Pred. No. 2.7e-40;
  Matches   99;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         381 QNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGL 440
               ||||||||||||||||||||||| | ||||| |:||||||| ||| ||||:||||  |||
Db           1 QNRRALDLLTAKRGGTCLFLGEECCXYVNQSGIITEKVKEIXDRIXCRAEDLQNTAPWGL 60

Qy         441 LSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
               ||||||| |||||||||:| |||||||||| |||||||||||||||:||||||
Db          61 LSQWMPWTLPFLGPLAAIIFLLLFGPCIFNFLVKFVSSRIEAVKLQIVLQMEP 113


RESULT 15
US-10-114-104-109
; Sequence 109, Application US/10114104
; Publication No. US20030198647A1
;   GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/114,104
;             FILING DATE: 03-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/979,847
;             FILING DATE: 26-NOV-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 109:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 162 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 109:
US-10-114-104-109

  Query Match            19.6%;  Score 516;  DB 14;  Length 162;
  Best Local Similarity  87.6%;  Pred. No. 2.7e-40;
  Matches   99;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         381 QNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGL 440
               ||||||||||||||||||||||| | ||||| |:||||||| ||| ||||:||||  |||
Db           1 QNRRALDLLTAKRGGTCLFLGEECCXYVNQSGIITEKVKEIXDRIXCRAEDLQNTAPWGL 60

Qy         441 LSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
               ||||||| |||||||||:| |||||||||| |||||||||||||||:||||||
Db          61 LSQWMPWTLPFLGPLAAIIFLLLFGPCIFNFLVKFVSSRIEAVKLQIVLQMEP 113


RESULT 16
US-09-793-451-4
; Sequence 4, Application US/09793451
; Publication No. US20030157597A1
; GENERAL INFORMATION:
;  APPLICANT: Arthur B. Raitano
;  APPLICANT:  Daniel E.H. Afar
;  APPLICANT:  Gazelle S. Rastegar
;  APPLICANT:  Steve Chappell Mitchell
;  APPLICANT:  Rene S. Hubert
;  APPLICANT:  Pia M. Challita-Eid
;  APPLICANT:  Mary Faris
;  APPLICANT:  Aya Jakobovits
;  TITLE OF INVENTION: 103P2D6:  TISSUE SPECIFIC PROTEIN HIGHLY
;  TITLE OF INVENTION:  EXPRESSED IN VARIOUS CANCERS
;  FILE REFERENCE: 129.2USU2
;  CURRENT APPLICATION NUMBER: US/09/793,451
;  CURRENT FILING DATE:  2001-02-26
;  PRIOR APPLICATION NUMBER: 60/184,558
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/218,856
;  PRIOR FILING DATE: 2000-07-13
;  NUMBER OF SEQ ID NOS: 752
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 576
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-793-451-4

  Query Match            19.5%;  Score 514.5;  DB 10;  Length 576;
  Best Local Similarity  25.5%;  Pred. No. 2.4e-39;
  Matches  142;  Conservative   63;  Mismatches  123;  Indels  229;  Gaps   16;
```

```
Qy        140 HSTPSPYKGLVLSKLHETLR-THTRLVSLFNTTLTRLHEVSAQNPT---NCWMCLPLHFR 195
              |  | |      | || |:  ||: |              :| ||:   |||:|: |
Db         26 HPLPLP------SYLHHTINLTHSLL--------------AASNPSLVNNCWLCISLSSS 65

Qy        196 PYISIP-VPEQW------------------------------------------------ 206
               | ::| |   |
Db         66 AYTAVPAVQTDWATSPISLHLRTSFNSPHLYPPEELIYFLDRSSKTSPDISHQQAAALLR 125

Qy        207 ---NNFSTEINTTSVLVGPLVS-------------------------------------- 225
                  | |  ||:|  : ||| :
Db        126 TYLKNLSPYINSTPPIFGPLTTQTTIPVAAPLCISWQRPTGIPLGNLSPSRCSFTLHLRS 185

Qy        226 ------------NLEITHTSNLTCVKFSNTIDTTSSQCIR-------------------- 253
                           | ||   ::   |  |    :|: |:
Db        186 PTTNINETIGAFQLHITDKPSINTDKLKN---ISSNYCLGRHLPCISLHPWLSSPCSSDS 242

Qy        254 ------------------------------------------------------------ 253

Db        243 PPRPSSCLLIPSPENNSERLLVDTRRFLIHHENRTFPSTQLPHQSPLQPLTAAALAGSLG 302

Qy        254 -WV--TP---PTRI------VCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY- 300
               ||  ||   |: :       ||  |:||:||:| | ||  :    | | || | :
Db        303 VWVQDTPFSTPSHLFTLHLQFCLAQGLFFLCGSSTYMCLPANWTGTCTLVFLTPKIQFAN 362

Qy        301 -TEQDLYNHVVPKPHNKRVPILPFVIRAGVLG---RLGTGIGSITTSTQFYYKLSQEING 356
               ||:     : |    : :|::| ::  |:      |||||  |:||   :  || : :
Db        363 GTEELPVPLMTPTQQKRVIPLIPLMVGLGLSASTVALGTGIAGISTSVMTFRSLSNDFSA 422

Qy        357 DMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTE 416
               :  :: :|  || |::||||||||||| ||||||::|| |:|| || |:|:||| :| :
Db        423 SITDISQTLSVLQAQVDSLAAVVLQNRRGLDLLTAEKGGLCIFLNEECCFYLNQSGLVYD 482

Qy        417 KVKEIRDRIQCRAEELQNTER--WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVK 474
               :|:::|| |  | :  |     |  || || |||| : ||  : |||||||||| |: :
Db        483 NIKKLKDRAQKLANQASNYAEPPWA-LSNWMSWVLPIVSPLIPIFLLLLFGPCIFRLVSQ 541

Qy        475 FVSSRIEAVKLQMVLQM 491
              |: :||:|:    : ||
Db        542 FIQNRIQAITNHSIRQM 558

RESULT 17
US-10-283-722-4
; Sequence 4, Application US/10283722
; Publication No. US20030194407A1
; GENERAL INFORMATION:
;  APPLICANT: Arthur B. Raitano
;  APPLICANT:  Daniel E.H. Afar
;  APPLICANT:  Gazelle S. Rastegar
;  APPLICANT:  Steve Chappell Mitchell
;  APPLICANT:  Rene S. Hubert
;  APPLICANT:  Pia M. Challita-Eid
;  APPLICANT:  Mary Faris
;  APPLICANT:  Aya Jakobovits
;  TITLE OF INVENTION: 103P2D6:  TISSUE SPECIFIC PROTEIN HIGHLY
;  TITLE OF INVENTION:  EXPRESSED IN VARIOUS CANCERS
;  FILE REFERENCE: 129.2USU2
;  CURRENT APPLICATION NUMBER: US/10/283,722
;  CURRENT FILING DATE:  2003-02-03
;  PRIOR APPLICATION NUMBER: US/09/793,451
;  PRIOR FILING DATE: 2001-02-26
;  PRIOR APPLICATION NUMBER: 60/184,558
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/218,856
;  PRIOR FILING DATE: 2000-07-13
;  NUMBER OF SEQ ID NOS: 752
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 576
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-283-722-4
```

```
  Query Match             19.5%;  Score 514.5;  DB 14;  Length 576;
  Best Local Similarity   25.5%;  Pred. No. 2.4e-39;
  Matches  142;  Conservative   63;  Mismatches  123;  Indels  229;  Gaps   16;

Qy        140 HSTPSPYKGLVLSKLHETLR-THTRLVSLFNTTLTRLHEVSAQNPT---NCWMCLPLHFR 195
              |  | |      | || |:  ||: |              :| ||:   |||:|: |
Db         26 HPLPLP------SYLHHTINLTHSLL--------------AASNPSLVNNCWLCISLSSS 65

Qy        196 PYISIP-VPEQW------------------------------------------------ 206
               | ::| |   |
Db         66 AYTAVPAVQTDWATSPISLHLRTSFNSPHLYPPEELIYFLDRSSKTSPDISHQQAAALLR 125

Qy        207 ---NNFSTEINTTSVLVGPLVS-------------------------------------- 225
                  | |  ||:|  : ||| :
Db        126 TYLKNLSPYINSTPPIFGPLTTQTTIPVAAPLCISWQRPTGIPLGNLSPSRCSFTLHLRS 185

Qy        226 ------------NLEITHTSNLTCVKFSNTIDTTSSQCIR-------------------- 253
                           | ||   ::   |  |    :|: |:
Db        186 PTTNINETIGAFQLHITDKPSINTDKLKN---ISSNYCLGRHLPCISLHPWLSSPCSSDS 242

Qy        254 ------------------------------------------------------------ 253

Db        243 PPRPSSCLLIPSPENNSERLLVDTRRFLIHHENRTFPSTQLPHQSPLQPLTAAALAGSLG 302

Qy        254 -WV--TP---PTRI------VCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY- 300
               ||  ||   |: :       ||  |:||:||:| | ||  :    | | || | :
Db        303 VWVQDTPFSTPSHLFTLHLQFCLAQGLFFLCGSSTYMCLPANWTGTCTLVFLTPKIQFAN 362

Qy        301 -TEQDLYNHVVPKPHNKRVPILPFVIRAGVLG---RLGTGIGSITTSTQFYYKLSQEING 356
               ||:     : |    : :|::| ::  |:      |||||  |:||   :  || : :
Db        363 GTEELPVPLMTPTQQKRVIPLIPLMVGLGLSASTVALGTGIAGISTSVMTFRSLSNDFSA 422

Qy        357 DMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTE 416
               :  :: :|  || |::||||||||||| ||||||::|| |:|| || |:|:||| :| :
Db        423 SITDISQTLSVLQAQVDSLAAVVLQNRRGLDLLTAEKGGLCIFLNEECCFYLNQSGLVYD 482

Qy        417 KVKEIRDRIQCRAEELQNTER--WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVK 474
               :|:::|| |  | :  |     |  || || |||| : ||  : |||||||||| |: :
Db        483 NIKKLKDRAQKLANQASNYAEPPWA-LSNWMSWVLPIVSPLIPIFLLLLFGPCIFRLVSQ 541

Qy        475 FVSSRIEAVKLQMVLQM 491
              |: :||:|:    : ||
Db        542 FIQNRIQAITNHSIRQM 558

RESULT 18
US-10-283-903-4
; Sequence 4, Application US/10283903
; Publication No. US20030219766A1
; GENERAL INFORMATION:
;  APPLICANT: Arthur B. Raitano
;  APPLICANT:  Daniel E.H. Afar
;  APPLICANT:  Gazelle S. Rastegar
;  APPLICANT:  Steve Chappell Mitchell
;  APPLICANT:  Rene S. Hubert
;  APPLICANT:  Pia M. Challita-Eid
;  APPLICANT:  Mary Faris
;  APPLICANT:  Aya Jakobovits
;  TITLE OF INVENTION: 103P2D6:  TISSUE SPECIFIC PROTEIN HIGHLY
;  TITLE OF INVENTION:  EXPRESSED IN VARIOUS CANCERS
;  FILE REFERENCE: 129.2USU2
;  CURRENT APPLICATION NUMBER: US/10/283,903
;  CURRENT FILING DATE:  2003-02-03
;  PRIOR APPLICATION NUMBER: US/09/793,451
;  PRIOR FILING DATE: 2001-02-26
;  PRIOR APPLICATION NUMBER: 60/184,558
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/218,856
;  PRIOR FILING DATE: 2000-07-13
;  NUMBER OF SEQ ID NOS: 752
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 576
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-283-903-4

  Query Match             19.5%;  Score 514.5;  DB 14;  Length 576;
  Best Local Similarity   25.5%;  Pred. No. 2.4e-39;
  Matches  142;  Conservative   63;  Mismatches  123;  Indels  229;  Gaps   16;

Qy        140 HSTPSPYKGLVLSKLHETLR-THTRLVSLFNTTLTRLHEVSAQNPT---NCWMCLPLHFR 195
              |  | |      | || ||:  ||: |              :| ||:   |||:|: |
Db         26 HPLPLP------SYLHHTINLTHSLL--------------AASNPSLVNNCWLCISLSSS 65

Qy        196 PYISIP-VPEQW------------------------------------------------ 206
               | ::| |   |
Db         66 AYTAVPAVQTDWATSPISLHLRTSFNSPHLYPPEELIYFLDRSSKTSPDISHQQAAALLR 125

Qy        207 ---NNFSTEINTTSVLVGPLVS-------------------------------------- 225
                  | |  ||:|  : ||| :
Db        126 TYLKNLSPYINSTPPIFGPLTTQTTIPVAAPLCISWQRPTGIPLGNLSPSRCSFTLHLRS 185

Qy        226 ------------NLEITHTSNLTCVKFSNTIDTTSSQCIR-------------------- 253
                           | ||   ::   |  |    :|: |:
Db        186 PTTNINETIGAFQLHITDKPSINTDKLKN---ISSNYCLGRHLPCISLHPWLSSPCSSDS 242

Qy        254 ------------------------------------------------------------ 253

Db        243 PPRPSSCLLIPSPENNSERLLVDTRRFLIHHENRTFPSTQLPHQSPLQPLTAAALAGSLG 302

Qy        254 -WV--TP---PTRI------VCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY- 300
               ||  ||   |: :       ||  |:||:||:| | ||  :    | | || | :
Db        303 VWVQDTPFSTPSHLFTLHLQFCLAQGLFFLCGSSTYMCLPANWTGTCTLVFLTPKIQFAN 362

Qy        301 -TEQDLYNHVVPKPHNKRVPILPFVIRAGVLG---RLGTGIGSITTSTQFYYKLSQEING 356
               ||:     : |    : :|::| ::  |:      |||||  |:||   :  || : :
Db        363 GTEELPVPLMTPTQQKRVIPLIPLMVGLGLSASTVALGTGIAGISTSVMTFRSLSNDFSA 422

Qy        357 DMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTE 416
               :  :: :|  || |::||||||||||| ||||||::|| |:|| || |:|:||| :| :
Db        423 SITDISQTLSVLQAQVDSLAAVVLQNRRGLDLLTAEKGGLCIFLNEECCFYLNQSGLVYD 482

Qy        417 KVKEIRDRIQCRAEELQNTER--WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVK 474
               :|:::|| |  | :  |     |  || || |||| : ||  : |||||||||| |: :
Db        483 NIKKLKDRAQKLANQASNYAEPPWA-LSNWMSWVLPIVSPLIPIFLLLLFGPCIFRLVSQ 541

Qy        475 FVSSRIEAVKLQMVLQM 491
              |: :||:|:    : ||
Db        542 FIQNRIQAITNHSIRQM 558

RESULT 19
US-10-001-835-168
; Sequence 168, Application US/10001835
; Publication No. US20020160387A1
; GENERAL INFORMATION:
;  APPLICANT: Salceda, Susana
;  APPLICANT:  Macina, Roberto
;  APPLICANT:  Recipon, Herve
;  APPLICANT:  Cafferkey, Robert
;  APPLICANT:  Sun, Yongming
;  APPLICANT:  Liu, Chenghua
;  TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
;  FILE REFERENCE: DEX-0277
;  CURRENT APPLICATION NUMBER: US/10/001,835
;  CURRENT FILING DATE:  2001-11-20
;  PRIOR APPLICATION NUMBER: 60/249,997
;  PRIOR FILING DATE: 2000-11-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 168
;   LENGTH: 469
;   TYPE: PRT
;   ORGANISM: Homo sapien
```

```
US-10-001-835-168

  Query Match             18.2%;  Score 480.5;  DB 13;  Length 469;
  Best Local Similarity   31.4%;  Pred. No. 3e-36;
  Matches  132;  Conservative   62;  Mismatches  141;  Indels   85;  Gaps   15;

Qy        136 LTRGHSTPSPYKGLVLSKLHETLR-----TH-TRLVSLFNTTLT--------RLHEVSAQ 181
              ::|   |  |   |  |:   ||      || |  :  |   :|        :|  :|:
Db         53 ISRQRPTGIPLGNLSPSRCSFTLHLRSPTTHITETIGAFQLHITDKPSINTDKLKNISSN 112

Qy        182 NPTNCWMCLPLHF-------------------RPYISIPVPEQWNNFSTEI--------- 213
                     ||  |                    ||   : :|   ||    :
Db        113 ------YCLGRHLPSISLHPWLPSPCSSDSPPRPSSRLLIPSPKNNSERLLVDTQRFLIH 166

Qy        214 --NTTSVLVGPLVSNLEITHTS---NLTCVKFSNTIDTTSSQCIRWV------TPPTRIV 262
                | ||        : :: | |    ||    : ::         ||      ||    :
Db        167 HENRTS-------PSTQLPHQSPLQPLTAASLAGSLGI-------WVQDTPFSTPHLFTL 212

Qy        263 ----CLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY--TEQDLYNHVVPKPHNK 316
                  ||  |:||:||:| | ||  :    | | || | :     ||:     : |    :
Db        213 HLQFCLTQGLFFLCGSSTYMCLPANWTGTCTLVFLTPKIQFANGTEELPVPLMTPTRQKR 272

Qy        317 RVPILPFVIRAGVLG---RLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLN 373
               :|::| ::  |:      |||||  |:||   :  || : :  :  :: :|  || |::
Db        273 VIPLIPLMVGLGLSASTIALGTGIAGISTSVTTFRSLSNDFSASITDISQTLSVLQAQVD 332

Qy        374 SLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQ 433
              ||||||||||| ||||||::|| |:|| || |:|:||| :| : :|:::|| |  | :
Db        333 SLAAVVLQNRRGLDLLTAEKGGLCIFLNEECCFYLNQSGLVYDNIKKLKDRAQKLANQAS 392

Qy        434 NTER--WGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQM 491
              |     |  ||  | |||| | ||  : ||||| |||| |: :|: :||:|:    : ||
Db        393 NYAEPPWA-LSNRMSWVLPILSPLIPIFLLLLFAPCIFCLVSQFIQNRIQAITNHSIAQM 451

RESULT 20
US-08-979-847-81
; Sequence 81, Application US/08979847
; Publication No. US20030039664A1
;  GENERAL INFORMATION:
;    APPLICANT:  PERRON, HERVE
;    APPLICANT:  BESEME, FREDERIC
;    APPLICANT:  BEDIN, FREDERIC
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  KOMURIAN-PRADEL, FLORENCE
;    APPLICANT:  JOLIVET-REYNAUD, COLETTE
;    APPLICANT:  MANDRAND, BERNARD
;    APPLICANT:  GARSON, JEREMY
;    APPLICANT:  TUKE, PHILIP
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  210
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/979,847
;      FILING DATE:  26-NOV-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERRIDGE, WILLIAM P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 39046A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  81:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  433 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-979-847-81

  Query Match             17.2%;  Score 452;  DB 8;  Length 433;
  Best Local Similarity   54.3%;  Pred. No. 1.3e-33;
  Matches   89;  Conservative   27;  Mismatches   38;  Indels   10;  Gaps    2;

Qy        334 GTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||||  ::||  :|: ||:  :  ::::  |::||| ||:||||: |||||   ||||::
Db          6 GTGIAGLSTSLSYYHTLSKNFSDSLQEIMKSILTLQSQLDSLAAMTLQNRRGPHLLTAEK 65

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI----QCRAEELQNTERWGLLSQWMPWVL 449
              || | ||||| |:| ||| || :    :::|     || :    |      |  |  |:|
Db         66 GGLCTFLGEECCFYTNQSGIVRDATWHLQERASDIRQCLSNSYTN------LWSWATWLL 119

Qy        450 PFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
              |||||:||::||| |||||| ||||||||||||:||||||||||
Db        120 PFLGPMAAILLLLTFGPCIFKLLVKFVSSRIEAIKLQMVLQMEP 163

RESULT 21
US-10-430-442-87
; Sequence 87, Application US/10430442
; Publication No. US20030186391A1
;  GENERAL INFORMATION:
;       APPLICANT: Herve PERRON
;                  Frederic BESEME
;                  Frederic BEDIN
;                  Glaucia PARANHOS-BACCALA
;                  Florence KOMURIAN-PRADEL
;                  Colette JOLIVET
;                  Bernard MANDRAND
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 92
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Oliff & Berridge
;            STREET: 700 South Washington Street, Suite 300
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: U.S.A.
;            ZIP: 22314
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/430,442
;            FILING DATE: 07-May-2003
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/691,563
;            FILING DATE: 02-AUG-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Berridge, William P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 38588
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 87:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 433 amino acids
```

```
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 87:
US-10-430-442-87

  Query Match             17.2%;  Score 452;  DB 14;  Length 433;
  Best Local Similarity   54.3%;  Pred. No. 1.3e-33;
  Matches   89;  Conservative   27;  Mismatches   38;  Indels   10;  Gaps    2;

Qy        334 GTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||||  ::||  :|: ||:  :  ::::  |::||| ||:||||: |||||   ||||::
Db          6 GTGIAGLSTSLSYYHTLSKNFSDSLQEIMKSILTLQSQLDSLAAMTLQNRRGPHLLTAEK 65

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI----QCRAEELQNTERWGLLSQWMPWVL 449
              || | ||||| |:| ||| || :    :::|     || :    |      |  |  |:|
Db         66 GGLCTFLGEECCFYTNQSGIVRDATWHLQERASDIRQCLSNSYTN------LWSWATWLL 119

Qy        450 PFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
              |||||:||::||| |||||| ||||||||||||:||||||||||
Db        120 PFLGPMAAILLLLTFGPCIFKLLVKFVSSRIEAIKLQMVLQMEP 163

RESULT 22
US-10-114-104-81
; Sequence 81, Application US/10114104
; Publication No. US20030198647A1
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/114,104
;            FILING DATE: 03-Apr-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/979,847
;            FILING DATE: 26-NOV-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 81:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 433 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: No. US20030198647A1 Relevant
;            TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 81:
US-10-114-104-81

  Query Match             17.2%;  Score 452;  DB 14;  Length 433;
  Best Local Similarity   54.3%;  Pred. No. 1.3e-33;
  Matches   89;  Conservative   27;  Mismatches   38;  Indels   10;  Gaps    2;

Qy        334 GTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
              ||||  ::||  :|: ||:  :  ::::  |::||| ||:||||: |||||   ||||::
Db          6 GTGIAGLSTSLSYYHTLSKNFSDSLQEIMKSILTLQSQLDSLAAMTLQNRRGPHLLTAEK 65

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI----QCRAEELQNTERWGLLSQWMPWVL 449
              || | ||||| |:| ||| || :    :::|     || :    |      |  |  |:|
Db         66 GGLCTFLGEECCFYTNQSGIVRDATWHLQERASDIRQCLSNSYTN------LWSWATWLL 119

Qy        450 PFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQMEP 493
              |||||:||::||| |||||| ||||||||||||:||||||||||
Db        120 PFLGPMAAILLLLTFGPCIFKLLVKFVSSRIEAIKLQMVLQMEP 163


RESULT 23
US-09-764-891-2970
; Sequence 2970, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2970
;   LENGTH: 253
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (162)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (195)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (206)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-2970

  Query Match             16.1%;  Score 423;  DB 10;  Length 253;
  Best Local Similarity   39.1%;  Pred. No. 3.4e-31;
  Matches   88;  Conservative   45;  Mismatches   72;  Indels   20;  Gaps    4;

Qy        275 SAYHCLNGSSESMCFLSFLVP-----------PMTIYTEQDLYNHVVPKPHNKR-VPILP 322
              | : ||  :    | : :: |           |: ||          | |  :| :  :|
Db         19 SIHQCLPSNWTGTCTIGYVTPDIFIAPGNLSLPIPIYGNS-------PLPRVRRAIHFIP 71

Qy        323 FVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQN 382
               :   |:|   ||||  || ::  | :||:||  ::: :  :| |:|:|::|||||||||
Db         72 LLAGLGILAGTGTGIAGITKASLTYSQLSKEIANNIDTMAKALTTMQEQIDSLAAVVLQN 131

Qy        383 RRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI-QCRAEELQNTERWGLL 441
              || ||:||| :|| || | |: |::||||  | : :::: ::    |    |    |
Db        132 RRGLDMLTAAQGGICLALDEKCCFWVNQSGXVQDNIRQLLNQASSLRERATQGWLNWEGT 191

Qy        442 SQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQ 486
               :|  ||||  ||| :|:||||||||: ||: :|||||::|:|||
Db        192 WKWXSWVLPLTGPLXSLLLLLLFGPCLLNLITQFVSSRLQAIKLQ 236


RESULT 24
US-10-276-774-1749
; Sequence 1749, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  APPLICANT:  Tang, Y, Tom et al
;  TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
;  FILE REFERENCE: 21272-030
;  CURRENT APPLICATION NUMBER: US/10/276,774
;  CURRENT FILING DATE:  2002-11-18
;  PRIOR APPLICATION NUMBER: 09/560,875
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  NUMBER OF SEQ ID NOS: 2700
;  SOFTWARE: Custom
; SEQ ID NO 1749
;   LENGTH: 618
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(618)
;   OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-1749

  Query Match             16.0%;  Score 421;  DB 15;  Length 618;
  Best Local Similarity   26.5%;  Pred. No. 1.9e-30;
  Matches  157;  Conservative   77;  Mismatches  214;  Indels  144;  Gaps   24;

Qy          6 HTFLFTVLLPPFAL---TAPPPCCCTTSSSPY----QEFLXRTRLPGNIDAPSYRSLSKG 58
              | ||| :|||   :    :|||   | | :|:       | ::: |  : :  :  ||
Db         16 HKFLF-LLLPSLLMGYSESPPP--ITDSWAPFISLTHHVLSQSQSP--LSSNCWICLSTH 70

Qy         59 NSTFTAHTHMPRN--CYNSATLCMHAN---THYWTGKMINPSCPGGLGATVCWTYFTHTS 113
                 |||   :| :   :  : : :| :     :     : |    |  |       :  |
Db         71 TQXFTA---LPADLLTWTQSNVSLHISYLAIPFLADSFLKPVLXPGNSAKHLSFKLSSLS 127

Qy        114 MSDGGGIQGQAREKQVKEAISQL-TRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTL 172
              |  |  :   |    :   :: : |   |:  |  |         | | |  :|     |
Db        128 MVSGRAV---ALLHLIASGLTSIQTNTASSKPPIWGY--------LSTQTSFISPPPLCL 176

Qy        173 TRLHEVSAQNPTNCWM------------------CLPLHFRPYISIPVPEQWNN---FS- 210
              :| :     || :  |                  | |    |   |     |     ||
Db        177 SRTY----PNPAHATMVGQVPQSLCGLIFTLRTPCRPSILHPNYKIISTSAWQKVLCFSG 232

Qy        211 -----TEINTT-----------------------SVLVGPLVSNLEITH----------- 231
                   | :: |                       | | ||   |: |
Db        233 SPTIHTSLHLTTGSSFLSFHPIPGFPAANSALYVSSLKGPPGKNVTIPSPVTGTXQPPHR 292

Qy        232 -TSNLTCVKFS-------NTIDTTSSQ----------------------CIRWVTPPTRI 261
               :: ||  | :       |::    ||                       : |:   |
Db        293 GSNRLTVDKDNFFLSPKPNSLHQLPSQTPYQALTGAALAGSYPIWENENTLSWLPTFTYN 352

Qy        262 VCLPS-GIFFVCGTSAYHCLNGSSESMCFLSF-------LVPPMTIYTEQDLYNHVVPKP 313
               || :  :||:| |: | ||  :    | | |       | |  ||    :      |
Db        353 FCLSTPSLFFLCDTNXYLCLPANWSGTCTLVFQAPTINILPPNQTILISVEASISSSPIR 412

Qy        314 HNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLN 373
              :   : ::  :   |:   |||||  ||||   |  |   ::  :| :  |: :|| ||:
Db        413 NKWALHLITLLTGLGITAALGTGIAGITTSITSYQTLFTTLSNTVEDMHTSITSLQRQLD 472

Qy        374 SLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEEL- 432
               |  |:||| | ||||| ::||||::| || |: ||:| ||   |: : |    || ||
Db        473 FLVGVILQNWRVLDLLTTEKGGTCIYLQEECCFCVNESGIVHIAVRRLHD----RAAELX 528

Qy        433 -QNTERW---GLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
               |  : |     | :|:||| ||||||  | |||: |||||||: :|:| |:
Db        529 HQVADSWWQGSSLLRWIPWVAPFLGPLIFLFLLLMIGPCIFNLVSRFISQRL 580


RESULT 25
US-09-954-983-11
; Sequence 11, Application US/09954983
; Publication No. US20020137889A1
; GENERAL INFORMATION:
;  APPLICANT: SOONG, NAY WEI
;  APPLICANT:  STEMMER, WILLEM P.
;  APPLICANT:  POWELL, SHARON K.
;  APPLICANT:  OTTO, EDWARD
;  TITLE OF INVENTION: STRESS RESISTANT RETROVIRUSES
;  FILE REFERENCE: 02-107120PC
;  CURRENT APPLICATION NUMBER: US/09/954,983
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: 60/233,398
;  PRIOR FILING DATE: 2000-09-18
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 676
;   TYPE: PRT
;   ORGANISM: Friend murine leukemia virus
US-09-954-983-11

  Query Match             14.9%;  Score 393;  DB 9;  Length 676;
  Best Local Similarity   23.9%;  Pred. No. 9.8e-28;
  Matches  143;  Conservative   81;  Mismatches  219;  Indels  156;  Gaps   22;

Qy           16 PFALTAPPPCCCTTS----------SSPYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF 62
                |::    ||||  :|          : |      |     |   :|  :::| |:|
Db           97 PYSSPPGPPCCSGSSGNSAGCSRDCNEPLTSLTPRCNTAWNRLKLDQVTHKS-SEGFYVC 155

Qy           63 TAHTHMPR---NC------YNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFT-HT 112
                   :| ||   :|      | ::  |      ||      ||          | | |
Db          156 PG-SHRPREAKSCGGPDSFYCASWGCETTGRVYW-----KPSS--------SWDYITVDN 201

Qy          113 SMSDGGGIQG----------QAREKQVKEAISQLTRGHS--------------------- 141
                :::    :|             |       ::  | |||
Db          202 NLTTNQAVQVCKDNKWCNPLAIRFTNAGRQVTSWTTGHSWGLRLYVTGKDPGLTFGIRLK 261

Qy          142 ---------------------------TPSPYKGLVLSKLHETL-------------RTH 161
                                            | | |    |    ||              |
Db          262 YQNLGPRVPIGPNPVLADQLSFPLPNPQPKPAKSPPASNSTPTLISPSPTPTQPPPAGTG 321

Qy          162 TRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVG 221
                 ||::|       |:  :      ||:||      |  : |        |  |  ||
Db          322 DRLLNLVQGAYQALNLTNPDKTQECWLCLVSGPPYYEGVAV------LGTYSNHTS---A 372

Qy          222 PLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC--------------LPSG 267
                |   :    |   |: |       |    ||  |    : :|               |:|
Db          373 PANCSAASQHKLTLSEV-------TGRGLCIGTVPKTHQALCNTTLKTGKGSYYLVAPAG 425

Qy          268 IFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI--- 320
                  : | |    ||: :    :   | |  | | :| :    :|:       :|| |:
Db          426 TMWACNTGLTPCLSATVLNRTTDYCVLVELWPRVTYHPPSYVYSQFENSYRHKREPVSLT 485

Qy          321 LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
                |  ::    :| :  |:|: ||   :|| : :|   :  |:::|  |:  |:  | ||:
Db          486 LALLLGGLTMGGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLTSLSE 545

Qy          378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
                ||||||| ||||  | || |  || |:| : : :| : : ::|:|:  |  :  ::::
Db          546 VVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLTQRQKLFESSQG 605

Qy          438 W--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
                |  || ::  ||    :   :|||  |:|:||||||| | ||:||  ||  |:  :: |
Db          606 WFEGLFNR-SPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFVKDRISVVQALVLTQ 663

RESULT 26
US-09-954-983-8
; Sequence 8, Application US/09954983
; Publication No. US20020137889A1
; GENERAL INFORMATION:
;  APPLICANT: SOONG, NAY WEI
;  APPLICANT:  STEMMER, WILLEM P.
;  APPLICANT:  POWELL, SHARON K.
```

```
;  APPLICANT:  OTTO, EDWARD
;  TITLE OF INVENTION: STRESS RESISTANT RETROVIRUSES
;  FILE REFERENCE: 02-107120PC
;  CURRENT APPLICATION NUMBER: US/09/954,983
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: 60/233,398
;  PRIOR FILING DATE: 2000-09-18
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 675
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Clone ID 2B-8
US-09-954-983-8

  Query Match             14.9%;  Score 392;  DB 9;  Length 675;
  Best Local Similarity   23.9%;  Pred. No. 1.2e-27;
  Matches  144;  Conservative   77;  Mismatches  218;  Indels  164;  Gaps   22;

Qy           16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF 62
                |::    ||||  :| :          |      |     |   :|  :::| |:|
Db           97 PYSSPPGPPCCSGSSGNSAGCFRDCDEPLTSLTPRCNTAWNRLKLDQVTHKS-SEGFYVC 155

Qy           63 TAHTHMPR---NC------YNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTS 113
                    | ||   :|      | ::  |      ||      ||          | | |
Db          156 PG-PHRPREAKSCGGPDSFYCASWGCETTGRVYW-----KPSS--------SWDYIT--- 198

Qy          114 MSDGGGIQGQA---------------REKQVKEAISQLTRGH------------------ 140
                  |      ||               |       ::  | ||
Db          199 -VDNNLTTNQAVKVCKDNKWCNPLAIRFTNTGRQVTSWTTGHYWGLRLYVTGKDPGLTFG 257

Qy          141 ------------------------------STPSPYKGLVLSKLHETL------------ 158
                                                | | |    |    ||
Db          258 IRLKYQNLGPRVPIGPNPVLADQLSFPLPNPQPRPAKSPPASNSTPTLISPSPTPTQPPP 317

Qy          159 -RTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTS 217
                  |  ||::|       |:  :      ||:||      |  : |        |  |  ||
Db          318 AGTGDRLLNLVQGAYQALNFTNPDKTQECWLCLVSGPPYYEGVAV------LGTYSNHTS 371

Qy          218 VLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC-------------- 263
                    |   :    |   |: |       |    ||  |    : :|
Db          372 ---APANCSAASQHKLTLSEV-------TGRGLCIGTVPKTHQALCNTTLKTGKGSYYLV 421

Qy          264 LPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVP 319
                 |:|  : | |    ||: :    :   | |: | | :| :    :|:       :|| |
Db          422 APAGTMWACNTGLTPCLSATVLNRTTDYCVLAELWPRVTYHPPSYVYSQFENSYRHKREP 481

Qy          320 I---LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLN 373
                :   |  ::    :| :  |:|: ||   :|| : :|   :  |:::|  |:  |:  |
Db          482 VSLTLALLLGGLTMGGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLT 541

Qy          374 SLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQ 433
                ||: ||||||| ||||  | || |  || |:| : : :| : : ::|:|:  |  :  :
Db          542 SLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFE 601

Qy          434 NTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQM 487
                :|: |  || ::  ||    :   :|||  |:::||||||| | ||:||  ||  |:  :
Db          602 STQGWFEGLFNR-SPWFTTLISTIMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALV 660

Qy          488 VLQ 490
                : |
Db          661 LTQ 663

RESULT 27
US-09-954-983-4
; Sequence 4, Application US/09954983
; Publication No. US20020137889A1
; GENERAL INFORMATION:
;  APPLICANT: SOONG, NAY WEI
; APPLICANT:  STEMMER, WILLEM P.
; APPLICANT:  POWELL, SHARON K.
; APPLICANT:  OTTO, EDWARD
; TITLE OF INVENTION: STRESS RESISTANT RETROVIRUSES
; FILE REFERENCE: 02-107120PC
; CURRENT APPLICATION NUMBER: US/09/954,983
; CURRENT FILING DATE:  2001-09-17
; PRIOR APPLICATION NUMBER: 60/233,398
; PRIOR FILING DATE: 2000-09-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;  LENGTH: 672
;  TYPE: PRT
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;  OTHER INFORMATION: Clone ID 2B-17
US-09-954-983-4

  Query Match             14.8%;  Score 391;  DB 9;  Length 672;
  Best Local Similarity   24.0%;  Pred. No. 1.5e-27;
  Matches  145;  Conservative   75;  Mismatches  219;  Indels  164;  Gaps   22;

Qy          16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF 62
               |::    ||||  :| :          |      |     |   :|  :::| |:|
Db          94 PYSSPPGPPCCSGSSGNSAGCFRDCDEPLTSLTPRCNTAWNRLKLDQVTHKS-SEGFYVC 152

Qy          63 TAHTHMPR---NC------YNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTS 113
                   | ||   :|      | ::  |      ||      ||          | | |
Db         153 PG-PHRPREAKSCGGPDSFYCASWGCETTGRVYW-----KPSS--------SWDYIT--- 195

Qy         114 MSDGGGIQGQA---------------REKQVKEAISQLTRGH------------------ 140
                 |      ||               |       ::  | ||
Db         196 -VDNNLTTNQAVKVCKDNKWCNPLAIRFTNTGRQVTSWTTGHYWGLRLYVTGKDPGLTFG 254

Qy         141 -----------------------------STPSPYKGLVLSKLHETL------------ 158
                                               | | |    |    ||
Db         255 IRLKYQNLGPRVPIGPNPVLADQLSFPLPNPQPKPAKSPPASNSTPTLISPSPTPTQPPP 314

Qy         159 -RTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTS 217
                 |  ||::|       |:  :      ||:||      |  : |        |  | ||
Db         315 AGTGDRLLNLVQGAYQALNLTNPDKTQECWLCLVSGPPYYEGVAV------LGTYSNHTS 368

Qy         218 VLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC-------------- 263
                   |   :    |   |: |       |    ||  |    : :|
Db         369 ---APANCSTASQHKLTLSEV-------TGRGLCIGTVPKTHQALCNTTLKTGKGSYYLV 418

Qy         264 LPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVP 319
                |:|  :  | |    ||: :    :   | |  | |:| :    :||       :||  |
Db         419 APAGTMWACNTGLTPCLSATVLNRTTDYCVLVKLWPRVTYHPPSYVYNQFENSYRHKREP 478

Qy         320 I---LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLN 373
               :   |  ::    :| :  |:|: ||   :|| : :|   :  |:::|  |:  |:  |
Db         479 VSLTLALLLGGLTMGGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLT 538

Qy         374 SLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQ 433
               ||: ||||||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  :
Db         539 SLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLTQRQKLFE 598

Qy         434 NTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQM 487
               ::: |  || ::  ||    :   :|||  |:|:||||||| | ||:||  ||  |:  :
Db         599 SSQGWFEGLFNR-SPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFVKDRISVVQALV 657

Qy         488 VLQ 490
               : |
Db         658 LTQ 660

RESULT 28
US-09-954-983-6
; Sequence 6, Application US/09954983
; Publication No. US20020137889A1
; GENERAL INFORMATION:
;  APPLICANT: SOONG, NAY WEI
;  APPLICANT:  STEMMER, WILLEM P.
;  APPLICANT:  POWELL, SHARON K.
;  APPLICANT:  OTTO, EDWARD
;  TITLE OF INVENTION: STRESS RESISTANT RETROVIRUSES
;  FILE REFERENCE: 02-107120PC
;  CURRENT APPLICATION NUMBER: US/09/954,983
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: 60/233,398
;  PRIOR FILING DATE: 2000-09-18
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 675
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Clone ID 4-7
US-09-954-983-6

  Query Match             14.8%;  Score 389;  DB 9;  Length 675;
  Best Local Similarity   24.0%;  Pred. No. 2.3e-27;
  Matches  145;  Conservative   75;  Mismatches  219;  Indels  164;  Gaps   22;

Qy          16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF 62
               |::    ||||  :| :          |      |     |   :|  ::|| |:|
Db          97 PYSSPPGPPCCSGSSGNSAGCFRDCDEPLTSLTPRCNTAWNRLKLDQVTHRS-SEGFYVC 155

Qy          63 TAHTHMPR---NC------YNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTS 113
                   | ||   :|      | ::  |      ||      ||          | | |
Db         156 PG-PHRPREAKSCGGPDSFYCASWGCETTGRVYW-----KPSS--------SWDYIT--- 198

Qy         114 MSDGGGIQGQA---------------REKQVKEAISQLTRGH------------------ 140
                 |      ||               |       ::  | ||
Db         199 -VDNNLTTNQAAQVCKDNKWCNPLAIRFTNAGRQVTSWTTGHYWGLRLYVTGKDPGLTFG 257

Qy         141 -----------------------------STPSPYKGLVLSKLHETL------------ 158
                                               | | |    |    ||
Db         258 IRLKYQNLGPRVPIGPNPVLADQLSFPLPNPLPKPAKSPPASNSTPTLISPSPTPTQPPP 317

Qy         159 -RTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTS 217
                 |  ||::|       |:  :      ||:||      |  : |        |  | ||
Db         318 AGTGDRLLNLVQGAYQALNLTNPDKTQECWLCLVSGPPYYEGVAV------LGTYSNHTS 371

Qy         218 VLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC-------------- 263
                   |   :    |   |: |       |    ||  |    : :|
Db         372 ---APANCSTASQHKLTLSEV-------TGRGLCIGTVPKTHQALCNTTLKAGKGSYYLV 421

Qy         264 LPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVP 319
                |:|  :  | |    ||: :    :   | |  | |:| :    :|:       :||  |
Db         422 APAGTMWACNTGLTPCLSATVLNRTTDYCVLVELWPRVTYHPPSYVYSQFENSYRHKREP 481

Qy         320 I---LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLN 373
               :   |  ::    :| :  |:|: ||   :|| : :|   :  |:::|  |:  |:  |
Db         482 VSLTLALLLGGLTMGGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLT 541

Qy         374 SLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQ 433
               ||: ||||||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  :
Db         542 SLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLTQRQKLFE 601

Qy         434 NTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQM 487
               ::: |  || ::  ||    :   :|||  |:|:||||||| | ||:||  ||  |:  :
Db         602 SSQGWFEGLFNR-SPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFVKDRISVVQALV 660

Qy         488 VLQ 490
               : |
Db         661 LTQ 663

RESULT 29
US-09-954-983-5
```

```
; Sequence 5, Application US/09954983
; Publication No. US20020137889A1
; GENERAL INFORMATION:
;  APPLICANT: SOONG, NAY WEI
;  APPLICANT:  STEMMER, WILLEM P.
;  APPLICANT:  POWELL, SHARON K.
;  APPLICANT:  OTTO, EDWARD
;  TITLE OF INVENTION: STRESS RESISTANT RETROVIRUSES
;  FILE REFERENCE: 02-107120PC
;  CURRENT APPLICATION NUMBER: US/09/954,983
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: 60/233,398
;  PRIOR FILING DATE: 2000-09-18
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 672
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Clone ID 4-4
US-09-954-983-5

  Query Match             14.7%;  Score 386;  DB 9;  Length 672;
  Best Local Similarity   23.9%;  Pred. No. 4.5e-27;
  Matches  144;  Conservative   76;  Mismatches  219;  Indels  164;  Gaps   22;

Qy         16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF 62
              |::    ||||  :| :          |      |     |   :|  :::| |:|
Db         94 PYSSPPGPPCCSGSSGNSAGCFRDCDEPLTSLTPRCNTAWNRLKLDQVTHKS-SEGFYVC 152

Qy         63 TAHTHMPR---NC------YNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTS 113
                  | ||   :|      | ::  |      ||      ||          | | |
Db        153 PG-PHRPREAKSCGGPDSFYCASWGCETTGRVYW-----KPSS--------SWDYIT--- 195

Qy        114 MSDGGGIQGQA---------------REKQVKEAISQLTRGH------------------ 140
                |      ||               |       ::  | ||
Db        196 -VDNNLTTNQAVKVCKDNKWCNPLAIRFTNTGRQVTSWTTGHYWGLRLYVTGKDPGLTFG 254

Qy        141 -----------------------------STPSPYKGLVLSKLHETL------------ 158
                                              | | |    |    ||
Db        255 IRLKYQNLGPRVPIGPNPVLADQLSFPLPNPLPKPAKSPPASNSTPTLISPSPTPTQPPP 314

Qy        159 -RTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTS 217
                |  ||::|       |:  :      ||:||      |  : |        |  | ||
Db        315 AGTGDRLLNLVQGAYQALNLTNPDKTQECWLCLVSGPPYYEGVAV------LGTYSNHTS 368

Qy        218 VLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC-------------- 263
                  |   :    |   |: |       |    ||  |    : :|
Db        369 ---APANCSTASQHKLTLSEV-------TGRGLCIGTVPKTHQALCNTTLKTGKGSYYLV 418

Qy        264 LPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVP 319
               |:|  : | |    ||: :    :   | |  | | :| :    :|:       :|| |
Db        419 APAGTMWACNTGLTPCLSATVLNRTTDYCVLVELWPRVTYHPPSYVYSQFENSYRHKREP 478

Qy        320 I---LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLN 373
              :   |  ::    :| :  |:|: ||   :|| : :|   :  |:::|  |:  |:  |
Db        479 VSLTLALLLGGLTMGGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLT 538

Qy        374 SLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQ 433
              ||: ||||||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  :
Db        539 SLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLTQRQKLFE 598

Qy        434 NTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQM 487
              ::: |  || ::  ||    :   :|||  |:|:||||||| | ||:||  ||  |:  :
Db        599 SSQGWFEGLFNR-SPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFVKDRISVVQALV 657

Qy        488 VLQ 490
              : |
Db        658 LTQ 660
```

```
RESULT 30
US-09-954-983-7
; Sequence 7, Application US/09954983
; Publication No. US20020137889A1
; GENERAL INFORMATION:
;  APPLICANT: SOONG, NAY WEI
;  APPLICANT:  STEMMER, WILLEM P.
;  APPLICANT:  POWELL, SHARON K.
;  APPLICANT:  OTTO, EDWARD
;  TITLE OF INVENTION: STRESS RESISTANT RETROVIRUSES
;  FILE REFERENCE: 02-107120PC
;  CURRENT APPLICATION NUMBER: US/09/954,983
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: 60/233,398
;  PRIOR FILING DATE: 2000-09-18
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 675
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Clone ID 2B-13
US-09-954-983-7

  Query Match             14.7%;  Score 386;  DB 9;  Length 675;
  Best Local Similarity   23.9%;  Pred. No. 4.5e-27;
  Matches  144;  Conservative   76;  Mismatches  219;  Indels  164;  Gaps   22;

Qy         16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF 62
              |::    ||||  :| :          |      |     |   :|  :::| |:|
Db         97 PYSSPPGPPCCSGSSGNSAGCFRDCDEPLTSLTPRCNTAWNRLKLDQVTHKS-SEGFYVC 155

Qy         63 TAHTHMPR---NC------YNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTS 113
                  | ||   :|      | ::  |      ||      ||          | | |
Db        156 PG-PHRPREAKSCGGPDSFYCASWGCETTGRVYW-----EPSS--------SWDYIT--- 198

Qy        114 MSDGGGIQGQA---------------REKQVKEAISQLTRGH------------------ 140
                |      ||               |       ::  | ||
Db        199 -VDNNLTTNQAVKVCKDNKWCNPLAIRFTNTGRQVTSWTTGHYWGLRLYVTGKDPGLTFG 257

Qy        141 -----------------------------STPSPYKGLVLSKLHETL------------ 158
                                              | | |    |    ||
Db        258 IRLKYQNLGPRVPIGPNPVLADQLSFPLPNPQPKPAKSPPASNSTPTLISPSPTPTQPPP 317

Qy        159 -RTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTS 217
                |  ||::|       |:  :      ||:||      |  : |        |  | ||
Db        318 AGTGDRLLNLVQGAYQALNFTNPDKTQECWLCLVSGPPYYEGVAV------LGTYSNHTS 371

Qy        218 VLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC-------------- 263
                  |   :    |   |: |       |    ||  |    : :|
Db        372 ---APANCSTASQHKLTLSEV-------TGRGLCIGTVPKTHQALCNTTLKTGKGSYYLV 421

Qy        264 LPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVP 319
               |:|  : | |    ||: :    :   | |  | | :| :    :|:       :|| |
Db        422 APAGTMWACNTGLTPCLSATVLNRTTDYCVLVELWPRVTYHPPSYVYSQFENSYRHKREP 481

Qy        320 I---LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLN 373
              :   |  ::    :| :  |:|: ||   :|| : :|   :  |:::|  |:  |:  |
Db        482 VSLTLALLLGGLTMGGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLT 541

Qy        374 SLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQ 433
              ||: ||||||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  :
Db        542 SLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLTQRQKLFE 601

Qy        434 NTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQM 487
              ::: |  || ::  ||    :   :|||  |:|:||||||| | ||:||  ||  |:  :
Db        602 SSQGWFEGLFNR-SPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFVKDRISVVQALV 660

Qy        488 VLQ 490
              : |
Db        661 LTQ 663
```

```
Search completed: January 22, 2005, 22:17:05
Job time : 110 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.


OM protein - protein search, using sw model

Run on:         January 22, 2005, 21:17:11 ; Search time 39 Seconds
                                          (without alignments)
                                          1216.279 Million cell updates/sec

Title:          US-09-319-156B-10
Perfect score:  2634
Sequence:       1 MALPYHTFLFTVLLPPFALT..........KFVSSRIEAVKLQMVLQMEP 493

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:      283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_79:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.

                                    SUMMARIES
                          %
Result                 Query
  No.     Score        Match Length DB  ID                    Description
 -------------------------------------------------------------------------------
     1      580    22.0    563  1  VCMVM7                env polyprotein -
     2    559.5    21.2    567  1  VCFVAS                env polyprotein -
     3      556    21.1    587  1  VCLJSA                env polyprotein -
     4      546    20.7    586  1  VCLJMP                env polyprotein -
     5    477.5    18.1    575  1  VCLJHD                env polyprotein pr
     6    476.5    18.1    570  4  B44282                retrovirus-related
     7    463.5    17.6    582  1  VCVDAR                env polyprotein -
     8    415.5    15.8    321  2  S12812                env polyprotein -
     9      409    15.5    676  2  T01381                env protein - muri
    10    407.5    15.5    665  1  VCVWEM                env polyprotein -
    11    402.5    15.3    636  1  VCVWFS                env polyprotein -
    12      400    15.2    688  2  A43491                env polyprotein -
    13      398    15.1    689  2  B43491                env polyprotein -
    14      393    14.9    676  1  VCMVPV                env polyprotein pr
    15    392.5    14.9    661  1  VCMVCB                env polyprotein -
    16      392    14.9    676  2  S70395                env polyprotein -
    17    391.5    14.9    662  1  VCVWGF                env polyprotein -
    18    391.5    14.9    662  2  A25982                env polyprotein -
    19    390.5    14.8    640  1  VCMVRV                env polyprotein pr
    20      390    14.8    669  1  VCVWEK                env polyprotein -
    21    389.5    14.8    662  1  VCMVLB                env polyprotein -
    22      386    14.7    627  1  VCVWM2                env polyprotein -
    23      385    14.6    668  1  VCMVFP                env polyprotein pr
    24      385    14.6    669  2  A46511                envelope protein -
    25      384    14.6    640  1  VCVWM1                env polyprotein -
    26      383    14.5    642  1  VCMVS2                env polyprotein pr
    27      383    14.5    645  1  VCMVSS                env polyprotein pr
    28      382    14.5    642  2  T10533                env polyprotein pr
    29    380.5    14.4    639  1  VCMVSA                env polyprotein pr
    30    380.5    14.4    666  1  VCMVHL                env polyprotein pr
    31      379    14.4    665  1  VCMVVR                env polyprotein pr
    32      377    14.3    665  1  VCMVKA                env polyprotein pr
    33      376    14.3    642  1  VCMVFG                env polyprotein -
    34      367    13.9    644  2  S15464                gp70 protein - mur
    35    364.5    13.8    353  2  S22805                env polyprotein -
    36      359    13.6    671  1  VCMVCE                env polyprotein -
    37    357.5    13.6    667  1  VCLJGL                env polyprotein pr
    38    311.5    11.8    201  2  A03985                env polyprotein -
    39    311.5    11.8    211  2  A03986                env polyprotein -
    40      226     8.6    488  2  A45714                envelope glycoprot
    41      217     8.2    488  1  VCLJCN                env polyprotein -
    42      209     7.9    488  1  VCLJMT                env polyprotein -
    43      208     7.9    488  2  S14605                envelope glycoprot
    44      204     7.7    486  1  VCLJH2                env polyprotein -
    45      203     7.7    488  1  VCVWH                 env polyprotein -


                                   ALIGNMENTS


RESULT 1
VCMVM7
env polyprotein - baboon endogenous virus (strain M7)
N;Alternate names: coat polyprotein
N;Contains: coat protein gp70; coat protein p20E
C;Species: baboon endogenous virus
A;Note: host Papio sp. (baboon)
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C;Accession: JT0262
R;Kato, S.; Matsuo, K.; Nishimura, N.; Takahashi, N.; Takano, T.
Jpn. J. Genet. 62, 127-137, 1987
A;Title: The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome
A;Reference number: JT0260
A;Accession: JT0262
A;Molecule type: DNA
A;Residues: 1-563 <KAT>
A;Cross-references: UNIPROT:P10269; GB:M16550; NID:g509586; PIDN:AAA87333.1; PID:g332599
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
F;1-376/Product: coat protein gp70 #status predicted <CGP>
F;377-563/Product: coat protein p20E #status predicted <CPP>
F;437-449/Region: immunosuppressive peptide #status predicted
F;113,219,229,264,282,292,306,312,321,339,469/Binding site: carbohydrate (Asn) (covalent)

  Query Match             22.0%;  Score 580;  DB 1;  Length 563;
  Best Local Similarity   29.2%;  Pred. No. 1.5e-36;
  Matches  146;  Conservative   98;  Mismatches  190;  Indels   66;  Gaps   14;

Qy         28 TTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNSTFTAHTHMPRNCYNSATLCMHANTHYW 87
              |: | | ||       | |    ||:|        : |:    :|| |   |   |  |:
Db         80 TSPSGPLQE------CPCN----SYQS--------SVHS----SCYTSYQQCRSGNKTYY 117

Qy         88 --------------------TGKMINPSCPGGLGATVCWTYFTHTSMSDGGGIQGQAREK 127
                                  | |:|   | |  | ::||:      :|||||     | |
Db        118 TATLLKTQTGGTSDVQVLGSTNKLIQSPCNGIKGQSICWSTTAPIHVSDGGGPLDTTRIK 177

Qy        128 QVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCW 187
               |:  : :: :       |  | : |: : |    : ::: | |   |   :     :||
Db        178 SVQRKLEEIHKALYPELQYHPLAIPKVRDNLMVDAQTLNILNATYNLLLMSNTSLVDDCW 237

Qy        188 MCLPLHFRPYISIPVPEQWNNFSTEI--NTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              :|| |   |   : :|    :: |    | : ::: ||:  ::    || :|: || : :
Db        238 LCLKL--GPPTPLAIPNFLLSYVTRSSDNISCLIIPPLL--VQPMQFSNSSCL-FSPSYN 292

Qy        246 TTS---------SQCIRWVTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVP 295
              :|          | |   :|  |  :|  :|  |:|| : ||  |  :   :| |: |:|
Db        293 STEEIDLGHVAFSNCTS-ITNVTGPICAVNGSVFLCGNNMAYTYLPTNWTGLCVLATLLP 351

Qy        296 PMTIYTEQD-----LYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKL 350
               : |    :       :| : :|  : :  :| :   |:     ||   :  |   | ||
Db          352 DIDIIPGDEPVPIPAIDHFIYRP-KRAIQFIPLLAGLGITAAFTTGATGLGVSVTQYTKL 410

Qy          351 SQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQ 410
                | ::  |:: :: ::  ||||::||| ||||||| ||||||::|| || | |: |:|||:
Db          411 SNQLISDVQILSSTIQDLQDQVDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYVNK 470

Qy          411 SRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFN 470
                | || :|:| ::: :: | ::| :   |  |   :|::|||||||  |:|||  ||||||
Db          471 SGIVRDKIKTLQEELERRRKDLASNPLWTGLQGLLPYLLPFLGPLLTLLLLLTIGPCIFN 530

Qy          471 LLVKFVSSRIEAVKLQMVLQ 490
                 |  |:: ::  :   ::  |
Db          531 RLTAFINDKLNIIHAMVLTQ 550


RESULT 2
VCFVAS
env polyprotein - avian spleen necrosis virus
N;Alternate names: coat polyprotein
N;Contains: coat protein gp22; coat protein gp73
C;Species: avian spleen necrosis virus
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C;Accession: A38212
R;Kewalramani, V.N.; Panganiban, A.T.; Emerman, M.
J. Virol. 66, 3026-3031, 1992
A;Title: Spleen necrosis virus, an avian immunosuppressive retrovirus, shares a receptor
A;Reference number: A38212; MUID:92219390; PMID:1313915
A;Accession: A38212
A;Molecule type: DNA
A;Residues: 1-567 <KEW>
A;Cross-references: UNIPROT:P31796; GB:M87666
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
F;1-401/Domain: extracellular #status predicted <EXT>
F;1-397/Product: coat protein gp73 #status predicted <CP1>
F;363-379/Region: hydrophobic #status predicted
F;394-397/Region: cleavage processing #status predicted
F;398-567/Product: coat protein gp22 #status predicted <CP2>
F;402-418/Domain: transmembrane #status predicted <TM1>
F;419-567/Domain: intracellular #status predicted <INT>
F;245,274,306,328,335/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             21.2%;  Score 559.5;  DB 1;  Length 567;
  Best Local Similarity   31.9%;  Pred. No. 5.5e-35;
  Matches  153;  Conservative   75;  Mismatches  158;  Indels   93;  Gaps   19;

Qy           66 THMPRNCYNSATLCMHANTHYWT---------------GKMINPSCPGGLGATVCWTYFT 110
                | |   ||     |      |:|                |:|  || | :|  |||
Db          121 TQMHSTCYEKTQECTLLGKTYFTAILQKTKLGSYEDGPNKLIQASCTGTVGKPVCWDPVA 180

Qy          111 HTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPS-PYKGLVLSKLHETLRTHTRLVSLFN 169
                   :|||||     ||: |:| : :: | || ||  |  | | :        :| | |
Db          181 PVYVSDGGGPTDMIREESVRERLEEIIR-HSYPSVQYHPLALPR--------SRGVDLDP 231

Qy          170 TT---LTRLHEV-SAQNP---TNCWMCLPLHFRPYISIPVPEQWN-----NFSTEI---- 213
                 |   |   |:| :| ||    |||:|: |     |   :|   |     | |  :
Db          232 QTSDILEATHQVLNATNPKLAENCWLCMTL--GTPIPAAIPTNGNVTLDGNCSLSLPFGC 289

Qy          214 ----------------NTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTP 257
                                | | : ||         | :| | ::    :   :||    :
Db          290 NPPGSIDVSCYAGEADNRTGIPVG-------YVHFTNCTSIQ---EVTNETSQ----MGN 335

Qy          258 PTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP--- 313
                 ||: | | |  |||| : ||  |      :| |: :|| ::| : ::     :| |
Db          336 LTRL-CPPPGHVFVCGNNMAYTALPNKWIGLCILASIVPDISIISGEE----PIPLPSIE 390

Qy          314 -----HNKRVPILPFVIRAGVLGRL---GTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
                     | : |  :| ::  |: |     |||:|    |   |:||| ::  |:: :: ::
Db          391 YTARRHKRAVQFIPLLVGLGISGATLAGGTGLG---VSVHTYHKLSNQLIEDVQALSGTI 447

Qy          366 VTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI 425
                  ||||::||| ||||||| ||||||::|| || | |: |:| |:| || :|:::::: :
Db          448 NDLQDQIDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIRKLQEDL 507

Qy          426 QCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVK 484
                  |   | :   |  |: ::|::|| ||||  ||| |  ||||   | : :  :|:|||
Db          508 IERKRALYDNPLWSGLNGFLPYLLPLLGPLFGLILFLTLGPCIMKTLTRIIHDKIQAVK 566


RESULT 3
VCLJSA
env polyprotein - simian AIDS retrovirus SRV-1
N;Alternate names: coat polyprotein
N;Contains: coat protein gp20; coat protein gp70
C;Species: simian AIDS retrovirus SRV-1
A;Note: host Macaca mulatta (rhesus macaque)
C;Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 16-Jul-1999
C;Accession: A04000
R;Power, M.D.; Marx, P.A.; Bryant, M.L.; Gardner, M.B.; Barr, P.J.; Luciw, P.A.
Science 231, 1567-1572, 1986
A;Title: Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrom
A;Reference number: A94711; MUID:86151668; PMID:3006247
A;Accession: A04000
A;Molecule type: DNA
A;Residues: 1-587 <POW>
A;Cross-references: GB:M11841; NID:g334746; PIDN:AAA47733.1; PID:g334750
C;Comment: The env polyprotein contains coat protein gp70 and coat protein gp20; however,
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: capsid protein; coat protein; glycoprotein; polyprotein
F;120,237,266,271,277,280,295,308,322,328,340,358,488,585/Binding site: carbohydrate (Asn

  Query Match             21.1%;  Score 556;  DB 1;  Length 587;
  Best Local Similarity   29.4%;  Pred. No. 1.1e-34;
  Matches  160;  Conservative   84;  Mismatches  201;  Indels   99;  Gaps   17;

Qy           19 LTAPPPCCCTT-SSSPYQEFLXRTRL-------------------PGNIDAPSYRSLSKG 58
                :::||    ||  | | |  :     |                   |   ::  ||  |:
Db           49 VSSPPTNSLTTVSCSTYTAYSVTNSLKWQCVSTPTTASPTHIGSCPSQCNSQSYDSVH-- 106

Qy           59 NSTFTAHTHMPRNCYNSATLCMHANTHYWTGKMI---NPS-------------------- 95
                             |||    |   |  | |  ||   :||
Db          107 -----------ATCYNHYQQCTIGNKTYLTATMIRDKSPSSGDGNVPTILGNNQNLIIAG 155

Qy           96 CP-GGLGATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKL 154
                ||    |  |||       |||||| | : ||  | :   :| :       |  | | :
Db          156 CPENKKGQVVCWNSQPSVHMSDGGGPQDKVREIIVNKKFEELHKSLFPELSYHPLALPEA 215

Qy          155 H--ETLRTHT-----RLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWN 207
                   | :  ||      : || | :  |      |   :||:||    |    :|:   ::
Db          216 RGKEKIDAHTFDLLATVHSLLNVSSQR------QLAEDCWLCL----RSGDPVPLALPYD 265

Qy          208 NFSTEINT-------TSVLVGP--LVSNLEITHTSNLTCVKFSNTID--------TTSSQ 250
                | |   :|        | |: |  ||     ||:  |     :|: |        |  |
Db          266 NTSCSNSTFFFNCSNCSCLITPPFLVQPFNFTHSVCLYADYQNNSFDIDVGLAGFTNCSS 325

Qy          251 CIRWVTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQD----- 304
                 |  :: |:  :| |:   |||| : ||  |  :    | |: |:| : |    :
Db          326 YIN-ISKPSSPLCAPNSSVFVCGNNKAYTYLPTNWTGSCVLATLLPDIDIIPGSEPVPIP 384

Qy          305 LYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDS 364
                  :| : :|  : :  :| ||  |:   : ||   :  |   | ||| ::  |:: :: :
Db          385 AIDHFLGRP-KRAIQFIPLVIGLGITTAVSTGTAGLGVSLTQYTKLSHQLISDVQAISST 443

Qy          365 LVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDR 424
                :  ||||::||| ||||||| ||||||::|| || | |: |:| |:| || :|:| ::|
Db          444 IQDLQDQVDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDD 503

Qy          425 IQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVK 484
                :: | ::| :   |      :|:|:| ||||  |:|:| ||| ||| |: |:  :||:::
Db          504 LEKRRKQLIDNPFWTGFHGLLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQ 563

Qy          485 LQMV 488
```

```
                : :
Db         564 AKPI 567


RESULT 4
VCLJMP
env polyprotein - Mason-Pfizer monkey virus
N;Alternate names: coat polyprotein
N;Contains: coat protein gp20; coat protein gp70
C;Species: Mason-Pfizer monkey virus
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 09-Jul-2004
C;Accession: D25839
R;Sonigo, P.; Barker, C.; Hunter, E.; Wain-Hobson, S.
Cell 45, 375-385, 1986
A;Title: Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type
A;Reference number: A90878; MUID:86189951; PMID:2421920
A;Accession: D25839
A;Molecule type: DNA
A;Residues: 1-586 <SON>
A;Cross-references: UNIPROT:P07575; GB:M12349; NID:g334702; PIDN:AAA47712.1; PID:g33470
A;Experimental source: clone 6A
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: capsid protein; coat protein; glycoprotein; polyprotein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-394/Product: coat protein gp70 #status predicted <GP7>
F;395-586/Product: coat protein gp20 #status predicted <GP2>
F;120,237,264,276,291,304,318,324,339,357,487/Binding site: carbohydrate (Asn) (covalent

  Query Match             20.7%;  Score 546;  DB 1;  Length 586;
  Best Local Similarity   28.6%;  Pred. No. 6.2e-34;
  Matches  145;  Conservative   91;  Mismatches  209;  Indels   62;  Gaps   13;

Qy          26 CCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNSTFTAHTHMPRNCYNSATLCMHANTH 85
               | :| ::|    :     ||  :  || |:               :|||    |   |
Db          78 CVSTPTTPSNTHIG--SCPGECNTISYDSVH-------------ASCYNHYQQCNIGNKT 122

Qy          86 YWTG------------------------KMINPSCPGG-LGATVCWTYFTHTSMSDGGGIQ 121
               | |                         :|   || |  |  |||       :|||||  |
Db         123 YLTATITGDRTPAIGDGNVPTVLGTSHNLITAGCPNGKKGQVVCWNSRPSVHISDGGGPQ 182

Qy         122 GQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH--ETLRTHTRLVSLFNTTLTRLHEVS 179
                :||:  | :   :| |       |  | | :    | :  ||  : |  |  : |:
Db         183 DKARDIIVNKKFEELHRSLFPELSYHPLALPEARGKEKIDAHT--LDLLATVHSLLNASQ 240

Qy         180 AQNPTNCWMCLPLHFRPYISIPVPEQ-WNNFSTEINTTSVLVGP-LVSNLEITHTSNLTC 237
                    :||:||       :::|  :   :||:   | :  |  | ||     | :: |
Db         241 PSLAEDCWLCLQSGDPVPLALPYNDTLCSNFACLSNHSCPLTPPFLVQPFNFTDSNCLYA 300

Qy         238 VKFSNTID-----TTSSQC-----IRWVTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSES 286
                  :|: |      : : |     :   : |:  :| |:   |||| : ||  |  :
Db         301 HYQNNSFDIDVGLASFTNCSSYYNVSTASKPSNSLCAPNSSVFVCGNNKAYTYLPTNWTG 360

Qy         287 MCFLSFLVPPMTIYTEQD-----LYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSIT 341
                | |: |:| :  |    :       :| :  |   : : ::|  :  |:   :  ||   :
Db         361 SCVLATLLPDIDIIPGSEPVPIPAIDHFLGKA-KRAIQLIPLFVGLGITTAVSTGAAGLG 419

Qy         342 TSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLG 401
                |   | ||| ::  |:: :: ::  ||||::||| ||||||| ||||||::|| || |
Db         420 VSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAEVVLQNRRGLDLLTAEQGGICLALQ 479

Qy         402 EERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILL 461
               |: |:| |:| || :|:| ::| :: |  :| :   |     ::|:|:| ||||  |:|:
Db         480 EKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPFWTSFHGFLPYVMPLLGPLLCLLLV 539

Qy         462 LLFGPCIFNLLVKFVSSRIEAVKLQMV 488
               | ||| ||| |: |:  :||::: : :
Db         540 LSFGPIIFNKLMTFIKHQIESIQAKPI 566


RESULT 5
VCLJHD
```

```
env polyprotein precursor - squirrel monkey retrovirus SMRV-H
N;Alternate names: coat polyprotein
N;Contains: coat protein gp20; outer membrane protein
C;Species: squirrel monkey retrovirus SMRV-H
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 09-Jul-2004
C;Accession: D31827
R;Oda, T.; Ikeda, S.; Watanabe, S.; Hatsushika, M.; Akiyama, K.; Mitsunobu, F.
Virology 167, 468-476, 1988
A;Title: Molecular cloning, complete nucleotide sequence, and gene structure of the prov
A;Reference number: A31827; MUID:89073750; PMID:3201749
A;Accession: D31827
A;Molecule type: DNA
A;Residues: 1-575 <ODA>
A;Cross-references: UNIPROT:P21412; GB:M23385; NID:g332626; PIDN:AAA66455.1; PID:g332627
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: capsid protein; coat protein; glycoprotein; polyprotein; transmembrane prote
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-386/Product: outer membrane protein #status predicted <OMP>
F;387-575/Product: coat protein gp20 #status predicted <TMP>
F;387-403/Domain: transmembrane #status predicted <TM1>
F;447-481/Region: immunosuppressive peptide
F;518-534/Domain: transmembrane #status predicted <TM2>
F;126,239,266,271,302,316,322,349,479/Binding site: carbohydrate (Asn) (covalent) #statu

  Query Match             18.1%;  Score 477.5;  DB 1;  Length 575;
  Best Local Similarity   28.5%;  Pred. No. 1.1e-28;
  Matches  136;  Conservative   81;  Mismatches  168;  Indels   93;  Gaps   16;

Qy          66 THMPRNCYNSATLCMHANTHYWTGKM--------INP-------------SCPGGLGATV 104
               : |  :||:| : |   |  |:|  :         ||              | | :| :|
Db         109 SQMHSSCYSSFSQCTQGNNTYFTAILQRTKSTSETNPVTSGLQPHGVLQAGCDGTVGKSV 168

Qy         105 CWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRL 164
               ||       :||||| |   ||  |:: |  : :       |  |  ||         ::
Db         169 CWNQQAPIHVSDGGGPQDAVRELYVQKQIELVIQSQFPKLSYHPLARSKPRGP-DIDAQM 227

Qy         165 VSLFNTTLTRLHE-VSAQNPT---NCWMCL------PLHF-----------------RPY 197
               : : : |    |: ::  ||:   |||:||      || |                  |:
Db         228 LDILSAT----HQALNISNPSLAQNCWLCLNQGTSMPLAFPVNISSFNASQNNCTPSLPF 283

Qy         198 ISIPVPEQ-WNNF--STEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRW 254
                  |:| |:  |    : |:  : ||  |:|        : |   ||
Db         284 RVQPMPSQVYPCFFKGAQNNSFDIPVG--VANF-------VNCSSSSN------------ 322

Qy         255 VTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP 313
                   :  :|   |  |||| : |:  |  :    | |: |:| : | :  |     || |
Db         323 ---HSEALCPGPGQAFVCGNNLAFTALPANWTGSCVLAALLPDIDIISGDD----PVPIP 375

Qy         314 --------HNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
                         : | ::| ::  ||   : ||   :  : | | ||| ::  |:: :: ::
Db         376 TFDYIAGRQKRAVTLIPLLVGLGVSTAVATGTAGLGVAVQSYTKLSHQLINDVQALSSTI 435

Qy         366 VTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI 425
                 |||||:||| ||||||| ||||||::|| || | |  |:| |:| || :|:| ::: :
Db         436 NDLQDQLDSLAEVVLQNRRGLDLLTAEQGGICLALQERCCFYANKSGIVRDKIKNLQEDL 495

Qy         426 QCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAV 483
               : | : | :      |:  :|::||||||| |:||   | | |   :   :  ::  ::
Db         496 EKRRKALADNLFLTGLNGLLPYLLPFLGPLFAIILFFSFAPWILRRVTALIRDQLNSL 553


RESULT 6
B44282
retrovirus-related env polyprotein pseudogene - human
N;Alternate names: coat polyprotein
N;Contains: coat protein gp70; coat protein p20E
C;Species: Homo sapiens (man)
C;Date: 17-Feb-1994 #sequence_revision 04-Jan-1996 #text_change 09-Jul-2004
C;Accession: B44282
R;Hirose, Y.; Takamatsu, M.; Harada, F.
Virology 192, 52-61, 1993
```

```
A;Title: Presence of env genes in members of the RTVL-H family of human endogenous retr
A;Reference number: A44282; MUID:93297138; PMID:8517031
A;Accession: B44282
A;Status: conceptual translation of pseudogene
A;Molecule type: mRNA
A;Residues: 1-570 <HIR>
A;Cross-references: UNIPROT:Q9N2K0; GB:D10083
C;Genetics:
A;Gene: env
C;Keywords: coat protein; polyprotein; pseudogene
F;1-372/Domain: coat protein gp70 <CGP>
F;370/Region: opal stop codon
F;373-569/Domain: coat protein p20E <CPP>
F;436-448/Region: immunosuppressive peptide

  Query Match             18.1%;  Score 476.5;  DB 4;  Length 570;
  Best Local Similarity   34.7%;  Pred. No. 1.3e-28;
  Matches  119;  Conservative   58;  Mismatches  115;  Indels   51;  Gaps   13;

Qy          189 CLPLHFRPYIS----------------IPVPEQWNNFSTEINTTSVLVGPLVSN---LEI 229
                |: ||  |::                 || ||  |:    |:|   |:     |   :::
Db          221 CISLH--PWLPSPCSSDSPPRPSSRLLIPSPEN-NSERLLIDTQCFLIHHENQNSPSMQL 277

Qy          230 THTS---NLTCVKFSNTIDTTSSQCIRWV--TP---PTRI------VCLPSGIFFVCGTS 275
                 | |    ||    : ::         ||  ||   |: :       ||  |:||:||:|
Db          278 PHQSPLQPLTAAALAGSLGV-------WVQDTPFSTPSHLFTLHLQFCLTQGLFFLCGSS 330

Qy          276 AYHCLNGSSESMCFLSFLVPPMTI--YTEQDLYNHVVPKPHNKRVPILPFVIRAGVLG-- 331
                 | ||  :    | | || | :     ||:     : |    : :|::| ::| |:
Db          331 TYMCLPANWTGTCTLVFLTPKIQFANRTEELPVPLMTPTXQKRVLPLIPLMVRLGLSAST 390

Qy          332 -RLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLT 390
                  |||||  |:||   :  |  : :  :  :: :|  || |::||||||||| | |||||
Db          391 IALGTGIAGISTSVTTFCSLCNDFSASITDISQTLSVLQAQVDSLAAVVLQNCRGLDLLT 450

Qy          391 AKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER--WGLLSQWMPWV 448
                |::|| |:|| || |:|:||| :| : :|:::|| |  | :  |     |  || || ||
Db          451 AEKGGPCIFLNEECCFYLNQSGLVYDNIKKLKDRAQKLANQASNYAEPPWA-LSNWMSWV 509

Qy          449 LPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQM 491
                || | ||  : || || |||| |: : : :||:|:    : ||
Db          510 LPILSPLIPIFLLPLFEPCIFCLVSQLIQNRIQAITNLSIRQM 552

RESULT 7
VCVDAR
env polyprotein - avian reticuloendotheliosis virus
N;Alternate names: coat polyprotein
N;Contains: coat protein gp22; coat protein gp73
C;Species: avian reticuloendotheliosis virus
C;Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 09-Jul-2004
C;Accession: A03999
R;Wilhelmsen, K.C.; Eggleton, K.; Temin, H.M.
J. Virol. 52, 172-182, 1984
A;Title: Nucleotide acid sequences of the oncogene v-rel in reticuloendotheliosis virus
A;Reference number: A93003; MUID:85009850; PMID:6090694
A;Accession: A03999
A;Molecule type: DNA
A;Residues: 1-582 <WIL>
A;Cross-references: UNIPROT:P03399
A;Experimental source: strain A
A;Note: strain A is a helper virus of the strain T
C;Comment: Enzymatic cleavages of env polyprotein may yield mature proteins including c
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; polyprotein

  Query Match             17.6%;  Score 463.5;  DB 1;  Length 582;
  Best Local Similarity   28.3%;  Pred. No. 1.3e-27;
  Matches  150;  Conservative   75;  Mismatches  186;  Indels  119;  Gaps   20;

Qy            9 LFTVLLPPFALTAPPPCCCTT-----SSSPYQE-------------FLXRTRLPGNIDA 49
```

```
                :|   : |     | || | |      |: |::               | :|:|    |
Db           96 VFKPKIIPSVQEQPGPCECLTIATQMHSTCYEKAQECTLLGKTYFTAILQKTKLGSYEDG 155

Qy           50 PSYRSLSKGNSTFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYF 109
                |:                          |   : |  |   |: | ||      ||
Db          156 PN-------------------------KLLQASCTGIWETSMLGPRCP-----CVC---- 181

Qy          110 THTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPS-PYKGLVLSKLHETLRTHTRLVSLF 168
                      ||||   :   :   | : :: | || ||  |  | | :         | | |
Db          182 -----LDGGGPTDRF-GRICAEGLEEIIR-HSYPSVQYHPLALPR--------PRGVDLD 226

Qy          169 NTT---LTRLHEV-SAQNP---TNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVG 221
                  |   |   |:| :| ||    |||:|: |  :     | | :    |      |: :
Db          227 PQTSDILEATHQVLNATNPQLAENCWLCMTLGTQS----PQPSRRMAMS-----LSMEIA 277

Qy          222 PLVSNLEITHTSN---LTC-------------VKFSNTIDTTSSQCIRWVTPPTRIVCLP 265
                 | |    ||  |   | |             | |:|      |   | :    |   |
Db          278 VLASLSGATHRVNRCQLLCREADNRTGIPVGYVHFTNCTSIQESLTRRVIYEILRDYVLH 337

Qy          266 SGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP--------HNKR 317
                  ::      ||  |      :| |: :|| |:|   ::     :| |        | :
Db          338 RVMYLCVEQHAYTALPNKWIGLCILASIVPDMSIIPGEE----PIPLPSIEYTAGRHKRA 393

Qy          318 VPILPFVIRAGVLGRL---GTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNS 374
                |  :| ::  |: |     |||:|    |   |:||| ::  |:: :: ::  ||||::|
Db          394 VQFIPLLVGLGITGATLAGGTGLG---VSVHTYHKLSNQLIEDVQALSGTINDLQDQIDS 450

Qy          375 LAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQN 434
                || ||||||| ||||||::|| || | |: |:| |:| || :|:::::: :  |   | :
Db          451 LAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIRKLQEDLLARKRALYD 510

Qy          435 TERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVK 484
                   |  |: ::|::|| ||||  ||| |  ||||   | : :  :|:  |
Db          511 NPLWNGLNGFLPYLLPSLGPLFGLILFLTLGPCIRKTLTRIIHDKIQGSK 560

RESULT 8
S12812
env polyprotein - feline endogenous virus RD114 (fragment)
C;Species: feline endogenous virus RD114
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 20-Sep-1999
C;Accession: S12812
R;Moehring, R.
submitted to the EMBL Data Library, February 1990
A;Reference number: S12812
A;Accession: S12812
A;Molecule type: genomic RNA
A;Residues: 1-321 <MOE>
A;Cross-references: EMBL:X51930; NID:g61676; PIDN:CAA36196.1; PID:g61677
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: polyprotein

  Query Match             15.8%;  Score 415.5;  DB 2;  Length 321;
  Best Local Similarity   31.2%;  Pred. No. 2.8e-24;
  Matches   96;  Conservative   71;  Mismatches  122;  Indels   19;  Gaps    6;

Qy          198 ISIPVPEQWNNFSTEINTTSVLVGP--LVSNLEITHTSNLTCVKFSNT-------IDTTS 248
                ::|| |    : :  :   |  : |  ||  :: :::| |:    ::|       :  |:
Db            5 LAIPTPSLTYSLADSLANASCQIIPPLLVQPMQFSNSSCLSSPFINDTEQIDLGAVTFTN 64

Qy          249 SQCIRWVTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQD--- 304
                   :  |: |   :|  :|  |:|| : ||  |  :   :|  : |:| : |    :
Db           65 CTSVANVSSP---LCALNGSVFLCGNNMAYTYLPQNWTRLCVQASLLPDIDINPGDEPVP 121

Qy          305 --LYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVT 362
                    :| : :|  : |  :| :   |:     ||   :  |   | ||| ::  |:: ::
Db          122 IPAIDHYIHRP-KRAVQFIPLLAGLGITAAFTTGATGLGVSVTQYTKLSHQLISDVQVLS 180

Qy          363 DSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIR 422
                 ::  ||||::||| ||||||| ||||||::|| || | |: |:| |:| ||  |:: ::
Db          181 GTIQDLQDQVDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRNKIRTLQ 240

Qy          423 DRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEA 482
```

```
               : :| | | |     |  |  ::|::|| ||||  |:|:|  ||:|:  |: |:: |:
Db         241 EELQKRRESLATNPLWTGLQGFLPYLLPLLGPLLTLLLILTIGPCVFSRLMAFINDRLNV 300

Qy         483 VKLQMVLQ 490
               |   :: |
Db         301 VHAMVLAQ 308


RESULT 9
T01381
env protein - murine leukemia virus
C;Species: murine leukemia virus
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C;Accession: T01381
R;Tumas, K.M.; Poszgay, J.M.; Avidan, N.; Ksiazek, S.J.; Overmoyer, B.; Blank, K.J.; Pry
Virology 192, 587-595, 1993
A;Title: Loss of antigenic epitopes as the result of env gene recombination in retrovir
A;Reference number: Z14313; MUID:93134803; PMID:7678475
A;Accession: T01381
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-676 <TUM>
A;Cross-references: UNIPROT:O12374; EMBL:S53043; NID:g263651
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein

  Query Match             15.5%;  Score 409;  DB 2;  Length 676;
  Best Local Similarity   23.7%;  Pred. No. 2.3e-23;
  Matches  142;  Conservative   81;  Mismatches  219;  Indels  158;  Gaps   22;

Qy          16 PFALTAPPPCCCTTS----------SSPYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF 62
               |::    ||||  :|          : |      |    ||   :|  :::| |:|
Db          97 PYSSAPGPPCCSGSSGNIAGCARDCNEPLTSLTPRCNTAGNRLKLDQVTHKS-SEGFYVC 155

Qy          63 TAHTHMPR---NC------YNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFT-HT 112
                  :| ||   :|      | ::  |      ||      || |        | | |
Db         156 PG-SHRPREAKSCGGPDSFYCASWGCETTGRVYW-----KPSSP--------WDYITVDN 201

Qy         113 SMSDGGGIQGQAREK----------QVKEAISQLTRGH---------------------- 140
               :::    :|     |             : ::  | ||
Db         202 NLTSNQAVQVCKDNKWCNPLAIQFTNAGKQVTSWTTGHYWGLRLYVSGQDPGLTFGIRLS 261

Qy         141 -------------------------STPSPYKGLVLSKLHETL-------------RTH 161
                                           | | |    |    ||              |
Db         262 YQNLGPRIPIGPNPVLADQLSFPLPNPLPKPAKSPPASSSTPTLISPSPTPTQPPPVGTG 321

Qy         162 TRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVG 221
                ||::|       |:  :      ||:||      |  :  |        |  | ||
Db         322 DRLLNLVQGAYQALNLTNPDKTQECWLCLVSGPPYYEGVAV------LGTYSNHTSA--- 372

Qy         222 PLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC--------------LPSG 267
                  :|  :     ||  :      |    ||  |    : :|               |:|
Db         373 --PANCSVASQHKLTLSEV-----TGRGLCIGTVPKTHQALCNTTLKAGKGSYYLVAPTG 425

Qy         268 IFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI--- 320
                 : | |    ||: :    :   | |  | | :| :    :|:       :|| |:
Db         426 TMWACNTGLTPCLSATVLNRTTDYCVLVELWPRVTYHPPSYVYSQFEKSHRHKREPVSLT 485

Qy         321 LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
               |  ::    :| :  |:|: ||   :|| : :|   :  |:::|  |:  |:  | ||:
Db         486 LALLLGGLTMGGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLTSLSE 545

Qy         378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
               ||||||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  ::::
Db         546 VVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLSQRQKLFESSQG 605

Qy         438 WGLLSQWM---PW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
               |     |    ||    :   :|||  |:|:||||||| | ||:||  ||  |:  :: |
Db         606 W--FEGWFNRSPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFVKDRISVVQALVLTQ 663


RESULT 10
VCVWEM
env polyprotein - Moloney murine leukemia virus
N;Contains: knob protein gp70; R protein; spike protein p15E
C;Species: Moloney murine leukemia virus
C;Date: 01-Sep-1981 #sequence_revision 24-Sep-1981 #text_change 09-Jul-2004
C;Accession: A93265; A93235; A93848; A03983
R;Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.
Nature 293, 543-548, 1981
A;Title: Nucleotide sequence of Moloney murine leukaemia virus.
A;Reference number: A93265; MUID:82035843; PMID:6169994
A;Accession: A93265
A;Molecule type: genomic RNA
A;Residues: 1-665 <SHI>
A;Cross-references: UNIPROT:P03385; GB:J02255; GB:J02256; GB:J02257; GB:M76668; NID:g331
A;Experimental source: clone pMLV-1
R;Sutcliffe, J.G.; Shinnick, T.M.; Green, N.; Liu, F.T.; Niman, H.L.; Lerner, R.A.
Nature 287, 801-805, 1980
A;Title: Chemical synthesis of a polypeptide predicted from nucleotide sequence allows d
A;Reference number: A93235; MUID:81052384; PMID:6159543
A;Accession: A93235
A;Molecule type: genomic RNA
A;Residues: 496-665 <SUT1>
A;Experimental source: provirus
R;Sutcliffe, J.G.; Shinnick, T.M.; Verma, I.M.; Lerner, R.A.
Proc. Natl. Acad. Sci. U.S.A. 77, 3302-3306, 1980
A;Title: Nucleotide sequence of Moloney leukemia virus: 3' end reveals details of replic
A;Reference number: A93848; MUID:81013872; PMID:6251454
A;Accession: A93848
A;Molecule type: DNA
A;Residues: 484-662,'CEF' <SUT2>
A;Experimental source: provirus, clone pMLV-201
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;34-469/Product: knob protein gp70 #status predicted <KNB>
F;470-649/Product: spike protein p15E #status predicted <SPK>
F;650-665/Product: R protein #status predicted <RPT>
F;45,199,326,358,398,434/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             15.5%;  Score 407.5;  DB 1;  Length 665;
  Best Local Similarity   25.0%;  Pred. No. 3e-23;
  Matches  146;  Conservative   77;  Mismatches  227;  Indels  133;  Gaps   21;

Qy          16 PFALTAPPPCCCTTSS--------SPYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF-- 62
               ||:    ||||   ||         |      |     |   :|  :::|    |  |
Db          96 PFSSPPGPPCCSGGSSPGCSRDCEEPLTSLTPRCNTAWNRLKLDQTTHKS----NEGFYV 151

Qy          63 TAHTHMPR---NC------YNSATLCMHANTHYWTGK------MINPSCPGGLGATVC-- 105
                   | ||   :|      | :   |      ||          :| :        ||
Db         152 CPGPHRPRESKSCGGPDSFYCAYWGCETTGRAYWKPSSSWDFITVNNNLTSDQAVQVCKD 211

Qy         106 --------------------------WTYFTHTSMSDGG---GI---------------- 120
                                         |    : |  | |   ||
Db         212 NKWCNPLVIRFTDAGRRVTSWTTGHYWGLRLYVSGQDPGLTFGIRLRYQNLGPRVPIGPN 271

Qy         121 ------QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH-ETLRTHTRLVSLFNTTLT 173
                     |  :: | ||          |   |  |  ||        |  ||::| :
Db         272 PVLADQQPLSKPKPVKSP--------SVTKPPSGTPLSPTQLPPAGTENRLLNLVDGAYQ 323

Qy         174 RLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT--- 230
                |:  |      ||:||      |  : |   ::| ::     ||     ::  |:|
Db         324 ALNLTSPDKTQECWLCLVAGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQG 383

Qy         231 -------HTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGS 283
                       |    |    ||  |:|      |         |:|  : | |    |:: :
Db         384 LCIGAVPKTHQALC----NTTQTSSRGSYYLVA--------PTGTMWACSTGLTPCISTT 431

Qy         284 ----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTG 336
                   :   | |  | | :| ::   :|        :|| |:   |  ::    :| :  |
Db         432 ILNLTTDYCVLVELWPRVTYHSPSYVYGLFERSNRHKREPVSLTLALLLGGLTMGGIAAG 491

Qy         337 IGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
```

```
              ||: ||   :|| : :|   :  |: :|  |:  |:  | ||: ||||||| ||||  |
Db        492 IGTGTTALMATQQFQQLQAAVQDDLREVEKSISNLEKSLTSLSEVVLQNRRGLDLLFLKE 551

Qy        394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW---- 447
              || |  | || |:| : : :| : : ::|:|:  | :  ::|: |  || ::  ||
Db        552 GGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFESTQGWFEGLFNR-SPWFTTL 610

Qy        448 VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              :   :|||  |:::||||||| | ||:||  ||  |:  :: |
Db        611 ISTIMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALVLTQ 653

RESULT 11
VCVWFS
env polyprotein - mink cell focus-forming virus
N;Alternate names: coat polyprotein
N;Contains: knob protein gp70; R protein; spike protein p15E
C;Species: mink cell focus-forming virus
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 23-Aug-1997
C;Accession: A31668
R;Chattopadhyay, S.K.; Baroudy, B.M.; Holmes, K.L.; Fredrickson, T.N.; Lander, M.R.; Mor
Virology 168, 90-100, 1989
A;Title: Biologic and molecular genetic characteristics of a unique MCF virus that is hi
A;Reference number: A31668; MUID:89085614; PMID:2535909
A;Accession: A31668
A;Molecule type: DNA
A;Residues: 1-636 <CHA>
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-440/Product: knob protein gp70 #status predicted <GP7>
F;441-620/Product: spike protein p15E #status predicted <P1E>
F;621-636/Product: R protein #status predicted <RPT>
F;43,58,297,329,369,405/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             15.3%;  Score 402.5;  DB 1;  Length 636;
  Best Local Similarity   25.8%;  Pred. No. 6.8e-23;
  Matches  132;  Conservative   81;  Mismatches  188;  Indels  111;  Gaps   18;

Qy         54 SLSKGNSTFTAHTHMPRN---CYNSATLCMHANTHYWTGKMINPSCPGG----------- 99
              || :||:        |||   ||:|: :             |  : |||
Db        149 SLKRGNT--------PRNQGPCYDSSVV----------SSGIQGATPGGRCNPLVLEFTD 190

Qy        100 LGATVCW-------------------TYFTHTSMSDGGG----------IQGQ---AREK 127
               |    |                   | |:  |      |          | ||   :|
Db        191 AGKKASWDGPKVWGLRLYRSTGIDPVTRFSLTRQVLNIGPRLPIGPNPVITGQLPPSRPV 250

Qy        128 QVKEAISQLTRGHSTPSPYKGLVLSKL---HETLRTHTRLVSLFNTTLTRLHEVSAQNPT 184
              |::     | |    | |    :: :     :   |  ||::| :     |:  |
Db        251 QIR-----LPRPPQPPPPGAASIVPETAPPSQQPGTGDRLLNLVDGAYQALNLTSPDKTQ 305

Qy        185 NCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT----------HTSN 234
               ||:||      |  : |   ::| ::     ||     ::  |:|           |
Db        306 ECWLCLVAGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQGLCVGAVPKTHQ 365

Qy        235 LTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGS----SESMCFL 290
                |    ||   ||          :  :  |:|  : | |    ||: :    :   | |
Db        366 ALC----NTTQKTSD--------GSYYLAAPAGTIWACNTGLTPCLSTTVLNLTTDYCVL 413

Qy        291 SFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTGIGSITT---ST 344
                | |:| ::   :|    |    :| |:   |  ::    :| :  |:|: ||   :|
Db        414 VELWPKVTYHSPDYVYTQFEPGARFRREPVSLTLALLLGGLTMGGIAAGVGTGTTALVAT 473

Qy        345 QFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEER 404
              | : :|   :: |:::|  |:  |:  |  ||: | ||||| ||||  |  || |  | ||
Db        474 QQFQQLQAAVHNDLKEVEKSITNLEKSLTSLSEVALQNRRGLDLLFLKEGGLCAALKEEC 533

Qy        405 CYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW----VLPFLGPLAAL 458
              |:| : : :| : : ::|:|:  | :  :: : |  || ::  ||    :   :|||  |
Db        534 CFYADHTGLVRDSMAKLRERLNQRQKLFESGQGWFEGLFNR-SPWFTTLISTIMGPLIVL 592
```

```
Qy        459 ILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              :|:||||||| | ||:||  ||  |:  :: |
Db        593 LLILLFGPCILNRLVQFVKDRISVVQALVLTQ 624

RESULT 12
A43491
env polyprotein - T3651/B murine leukemia virus (strain T1223/B)
C;Species: T3651/B murine leukemia virus
C;Date: 12-Jan-1993 #sequence_revision 12-Jan-1993 #text_change 09-Jul-2004
C;Accession: A43491
R;Baylac-Kalabokias, H.; Astier-Gin, T.; Moynet, D.; Hernould, M.; Mamoun, R.; Legrand,
Virus Res. 18, 117-134, 1991
A;Title: A new leukemogenic retrovirus isolated from tumor cells derived from a radio-inc
A;Reference number: A43491; MUID:91253260; PMID:2042396
A;Accession: A43491
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-688 <BAY>
A;Cross-references: UNIPROT:Q9Q9A5; GB:X59002
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei

  Query Match             15.2%;  Score 400;  DB 2;  Length 688;
  Best Local Similarity   23.7%;  Pred. No. 1.2e-22;
  Matches  141;  Conservative   84;  Mismatches  215;  Indels  156;  Gaps   21;

Qy         16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGNIDAPSYRSLSKGNSTFTAH 65
              ||:    ||||  :| |          |   :  |     |        |||       |
Db        111 PFSPPPGPPCCSGSSDSTPGCSRDCEEPLTSYTPRCNTAWN-----RLKLSK-----VTH 160

Qy         66 THM--------------------PRNCYNSATLCMHANTHYW---------------TGK 90
               |                     | : | ::  |       |               |
Db        161 AHNEGFYVCPGPHRPRWARSCGGPESFYCASWGCETTGRASWKPSSSWDYITVSNNLTSD 220

Qy         91 MINPSCPG--------------GLGATV-----CWTYFTHTSMSDGGGIQGQAREKQVKE 131
                 |:| |              |  ||       |    : |  | | | |      ::
Db        221 QATPACKGNKWCNSLTIRFTSFGKQATSWVTGHWWGLRLYVSGHDPGLIFG------IRL 274

Qy        132 AISQL--------------TRGHSTPSPYKGLVLSKLHET-----------LRTHTRLVS 166
               |: |               |  | | | :    |    |                ||::
Db        275 KITDLGPRVPIGPNPVLSDRRPPSRPRPTRSPPPSNSTPTETPLTLPEPPPAGVENRLLN 334

Qy        167 LFNTTLTRLHEVSAQNPTNCWMCL----PLH--------FRPYISIP----VPEQWNNFS 210
              |       |:  |      ||:||    | :        :  : | |    |  |
Db        335 LVKGAYQALNLTSPDKTQECWLCLVSGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTL 394

Qy        211 TEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFF 270
              :|:    : :| :   |::            ||   ||:         :  :  |:|  :
Db        395 SEVTGQGLCIGAVPKTLQV----------LCNTTQKTSA--------GSYYLAAPTGTIW 436

Qy        271 VCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPF 323
               | |    |:: :    :   | |  | | :| ::   :|:    :   || |:   |
Db        437 ACSTGLTPCISTTILNLTTDYCVLVELWPRVTYHSPSYVYHQFERRAKYKREPVSLTLAL 496

Qy        324 VIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVL 380
              ::    :| :  |:|: ||   :|| : :|   :: |:::|  |:  |:  | ||: |||
Db        497 LLGGLTMGGIAAGVGTGTTALVATQQFQQLQAAMHDDLKEVEKSITNLEKSLTSLSEVVL 556

Qy        381 QNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW-- 438
              |||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  :: : |
Db        557 QNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLSQRQKLFESQQGWFE 616

Qy        439 GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              || ::  ||    :   :|||  |:|:||||||| | ||:|:  ||  |:  :: |
Db        617 GLFNK-SPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFIKDRISVVQALVLTQ 671

RESULT 13
B43491
env polyprotein - T3651/B murine leukemia virus
C;Species: T3651/B murine leukemia virus
C;Date: 12-Jan-1993 #sequence_revision 12-Jan-1993 #text_change 09-Jul-2004
```

```
C;Accession: B43491
R;Baylac-Kalabokias, H.; Astier-Gin, T.; Moynet, D.; Hernould, M.; Mamoun, R.; Legrand,
Virus Res. 18, 117-134, 1991
A;Title: A new leukemogenic retrovirus isolated from tumor cells derived from a radio-ir
A;Reference number: A43491; MUID:91253260; PMID:2042396
A;Accession: B43491
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-689 <BAY>
A;Cross-references: UNIPROT:Q9Q9A5; GB:X59001
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei

  Query Match             15.1%;  Score 398;  DB 2;  Length 689;
  Best Local Similarity   23.5%;  Pred. No. 1.7e-22;
  Matches  140;  Conservative   85;  Mismatches  215;  Indels  156;  Gaps   21;

Qy         16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGNIDAPSYRSLSKGNSTFTAH 65
              ||:    ||||  :: |          |   :  |     |        |||       |
Db        112 PFSPPPGPPCCSGSNDSTSGCSRDCEEPLTSYTPRCNTAWN-----RLKLSK-----VTH 161

Qy         66 THM--------------------PRNCYNSATLCMHANTHYW---------------TGK 90
               |                     | : | ::  |       |               |
Db        162 AHNEGFYVCPGPHRPRWARSCGGPESFYCASWGCETTGRASWKPSSSWDYITVSNNLTAD 221

Qy         91 MINPSCPG--------------GLGATV-----CWTYFTHTSMSDGGGIQGQAREKQVKE 131
                 |:| |              |  ||       |    : |  | | | |      ::
Db        222 QATPACKGNKWCNSLTIRFTSFGKQATSWVTGHWWGLRLYVSGHDPGLIFG------IRL 275

Qy        132 AISQL--------------TRGHSTPSPYKGLVLSKLHET-----------LRTHTRLVS 166
               |: |               |  | | | :    |    |                ||::
Db        276 KITDLGPRVPIGPNPVLSDRRPPSRPRPTRSPPSSNSTPTETPLTLPEPPPAGVENRLLN 335

Qy        167 LFNTTLTRLHEVSAQNPTNCWMCL----PLH--------FRPYISIP----VPEQWNNFS 210
              |       |:  |      ||:||    | :        :  : | |    |  |
Db        336 LVKGAYQALNLTSPDKTQECWLCLVSGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTL 395

Qy        211 TEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFF 270
              :|:    : :| :   |::            ||   ||:         :  :  |:|  :
Db        396 SEVTGQGLCIGAVPKTLQV----------LCNTTQKTST--------GSYYLAAPTGTIW 437

Qy        271 VCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPF 323
               | |    |:: :    :   | |  | | :| ::   :|:    :   || |:   |
Db        438 ACSTGLTPCISTTILNLTTDYCVLVELWPRVTYHSPSYVYHQFERRAKYKREPVSLTLAL 497

Qy        324 VIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVL 380
              ::    :| :  |:|: ||   :|| : :|   :: |:::|  |:  |:  | ||: |||
Db        498 LLGGLTMGGIAAGVGTGTTALVATQQFQQLQAAMHDDLKEVEKSITNLEKSLTSLSEVVL 557

Qy        381 QNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW-- 438
              |||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  :: : |
Db        558 QNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLSQRQKLFESQQGWFE 617

Qy        439 GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              || ::  ||    :   :|||  |:|:||||||| | ||:|:  ||  |:  :: |
Db        618 GLFNR-SPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFIKDRISVVQALVLTQ 672

RESULT 14
VCMVPV
env polyprotein precursor - Friend murine leukemia virus (strain PVC-211)
N;Alternate names: coat polyprotein
N;Contains: knob protein gp76; R protein; spike protein p15E
C;Species: Friend murine leukemia virus
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C;Accession: A38210; S35476
R;Masuda, M.; Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.
J. Virol. 66, 2798-2806, 1992
A;Title: Molecular characterization of a neuropathogenic and nonerythroleukemogenic vari
A;Reference number: A38210; MUID:92219364; PMID:1560524
A;Accession: A38210
A;Molecule type: DNA
A;Residues: 1-676 <SPE>
A;Cross-references: UNIPROT:P26803; GB:M93134; NID:g331898; PIDN:AAA46478.1; PID:g331901
R;Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.; Masuda, M.
Nucleic Acids Res. 20, 3249, 1992
A;Title: Complete nucleotide sequence of a neuropathogenic variant of Friend murine leuke
A;Reference number: S35474; MUID:92319660; PMID:1620621
A;Accession: S35476
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-676 <REM>
A;Cross-references: EMBL:M93134; NID:g331898; PIDN:AAA46478.1; PID:g331901
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F;1-35/Domain: signal sequence #status predicted <SIG>
F;36-676/Product: env polyprotein #status predicted <ENV>
F;36-619/Domain: extracellular #status predicted <EXT>
F;36-479/Product: knob protein gp76 #status predicted <KGP>
F;476-479/Region: cleavage processing #status predicted
F;480-659/Product: spike protein p15E #status predicted <SPP>
F;486-502/Region: hydrophobic #status predicted
F;620-636/Domain: transmembrane #status predicted <TM1>
F;637-676/Domain: intracellular #status predicted <INT>
F;660-676/Product: R protein #status predicted <RPT>
F;46,202,336,368,408,444/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             14.9%;  Score 393;  DB 1;  Length 676;
  Best Local Similarity   24.1%;  Pred. No. 3.9e-22;
  Matches  142;  Conservative   85;  Mismatches  226;  Indels  136;  Gaps   22;

Qy         16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF 62
              |::    ||||  :| :          |      |     |   :|  :::  | |
Db         97 PYSSPPGPPCCSGSSGNRAGCARDCDEPLTSLTPRCNTAWNRLKLDQVTHK--SSGGFYV 154

Qy         63 TAHTHMPR---NC------YNSATLCMHANTHYWTGK------MINPSCPGGLGATVC-- 105
                 :| ||   :|      | ::  |      ||          :: :      | ||
Db        155 CPGSHRPRKAKSCGGPDSFYCASWGCETTGRAYWKPSSSWDYITVDNNLTTNQAAQVCKD 214

Qy        106 --------------------------WTYFTHTSMSDGG---GIQGQAREKQVKEAI--- 133
                                        |    : |  | |   ||: : :    :  |
Db        215 NKWCNPLAIQFTNAGKQVTSWTIGHYWGLRLYVSGQDPGLTFGIRLKYQNLGPRVPIGPN 274

Qy        134 ----SQLT--RGHSTPSPYKGLVLSKLHETL-------------RTHTRLVSLFNTTLTR 174
                   ||:    :  | | |    |    ||              |  ||::|
Db        275 PVLADQLSFPLPNPLPKPAKSPSASNSTPTLISPSPAPTQPPPAGTGDRLLNLVQGAYQA 334

Qy        175 LHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNN---FSTEINTTSVLVGPLVSNLEITH 231
              |:  :      ||:||         :  |  :       |  | ||    |   :    |
Db        335 LNLTNPDKTQECWLCL---------VSAPPYYEGVAVLGTYSNHTS---APANCSAGSQH 382

Qy        232 TSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC--------------LPSGIFFVCGTSAY 277
                 |: |       |    ||  |    : :|               |:|  : | |
Db        383 KLTLSEV-------TGQGLCIGTVPKTHQALCNTTLKTGKGSYYLVAPAGTMWACNTGLT 435

Qy        278 HCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVL 330
               ||: :    :   | |  | | :| :    :|:       :|| |:   |  ::    :
Db        436 PCLSATVLNRTTDYCVLVELWPRVTYHPPSYVYSQFEKSYRHKREPVSLTLALLLGGLTM 495

Qy        331 GRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALD 387
              | :  |:|: ||   :|| : :|   :  |:::|  |:  |:  | ||: ||||||| ||
Db        496 GGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLTSLSEVVLQNRRGLD 555

Qy        388 LLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWM 445
              ||  | || |  | || |:| : : :| : : ::|:|:  | :  :::: |  || ::
Db        556 LLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLTQRQKLFESSQGWFEGLFNR-S 614

Qy        446 PW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              ||    :   :|||  |:|:||||||| | ||:||  ||  |:  :: |
Db        615 PWFTTLISTIMGPLIILLLILLFGPCILNRLVQFVKDRISVVQALVLTQ 663

RESULT 15
```

```
VCMVCB
env polyprotein - Cas-Br-E murine leukemia virus
N;Contains: coat protein p15E; knob protein gp70 precursor
C;Species: Cas-Br-E murine leukemia virus
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 09-Jul-2004
C;Accession: B26103
R;Rassart, E.; Nelbach, L.; Jolicoeur, P.
J. Virol. 60, 910-919, 1986
A;Title: Cas-Br-E murine leukemia virus: sequencing of the paralytogenic region of its g
kemic tissues.
A;Reference number: A26103; MUID:87061215; PMID:3023680
A;Accession: B26103
A;Molecule type: DNA
A;Residues: 1-661 <RAS>
A;Cross-references: UNIPROT:P08360; GB:M14702; NID:g332016; PIDN:AAA46512.1; PID:g332018
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-465/Product: knob protein gp70 #status predicted <KOP>
F;466-661/Product: coat protein p15E #status predicted <PTE>
F;43,186,199,322,361,394,430,598/Binding site: carbohydrate (Asn) (covalent) #status pre

  Query Match            14.9%;  Score 392.5;  DB 1;  Length 661;
  Best Local Similarity  22.8%;  Pred. No. 4.2e-22;
  Matches  142;  Conservative   76;  Mismatches  193;  Indels  211;  Gaps   23;

Qy          15 PPFALTAPPPCC---------CT-------TSSSPYQEFLXRTRLPGNIDAPSYRSLSKG 58
               ||::    ||||         |        || ||                   |  :
Db          93 PPYSSPPGPPCCSGDAGAVSGCARDCDEPLTSYSP-------------------RCNTAW 133

Qy          59 NSTFTAH-THMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVC----------- 105
               |    |  || |:  :    :|  ::   |       || ||| |  |
Db         134 NRLKLARVTHAPKEGF---YICPGSHRPRWA-----RSC-GGLDAYYCASWGCETTGRAA 184

Qy         106 ------WTYFTHTSMSDGGGIQGQAREKQVKEA--------------------ISQLTRG 139
                     | | |         :       |  :|                     :  |  |
Db         185 WNPTSSWDYIT---------VSNNLTSSQATKACKNNGWCNPLVIRFTGPGKRATSWTTG 235

Qy         140 H----------------------------------------------------------- 140
               |
Db         236 HFWGLRLYISGHDPGLTFGIRLKVTDLGPRVPIGPNPVLSDQRPPSRPVPARPPPPSASP 295

Qy         141 STPS--PYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYI 198
               |||:  | :|           |  ||::|       |:         ||:||      |
Db         296 STPTIPPQQG-----------TGDRLLNLVQGAYLTLNMTDPTRTQECWLCLVSEPPYYE 344

Qy         199 SIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWV--- 255
                : |  :: :  |          |   :    |   |: |       |   :|:  |
Db         345 GVAVLREYTSHET---------APANCSSGSQHKLTLSEV-------TGQGRCLGTVPKT 388

Qy         256 -------TPPT----RIVCLPSGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIY 300
                      | ||      :  | |  : | |    ||: :    :   | |  | | :| :
Db         389 HQALCNRTEPTVSGSNYLVAPEGTLWACSTGLTPCLSTTVLNLTTDYCVLVELWPKVTYH 448

Qy         301 TEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEI 354
               :   :|    |    :| |:   |  :     :| :  |:|: ||   :|| : :|   :
Db         449 SPDYVYTQFEPGARFRREPVSLTLALLPEGLTMGGIAAGVGTGTTALVATQQFQQLQAAM 508

Qy         355 NGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIV 414
               : |:::|  |:  |:  | ||: ||||||| ||||  | || |  | || |:| : : :|
Db         509 HNDLKEVEKSITNLEKSLTSLSEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLV 568

Qy         415 TEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCI 468
                : :: |:|:  | :  :: : |  || ::  ||    :   :|||  |:|:|||||||
Db         569 RDSMAKLRERLNQRQKLFESGQGWFEGLFNR-SPWFTTLISTIMGPLIVLLLILLFGPCI 627

Qy         469 FNLLVKFVSSRIEAVKLQMVLQ 490
                | ||:||  ||  |:  :: |
Db         628 LNRLVQFVKDRISVVQALVLTQ 649
```

```
RESULT 16
S70395
env polyprotein - Friend murine leukemia virus (strain FB29)
N;Contains: knob protein gp76; R protein; spike protein p15E
C;Species: Friend murine leukemia virus
A;Variety: strain FB29
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S70395
R;Perryman, S.; Nishio, J.; Chesebro, B.
Nucleic Acids Res. 19, 6950, 1991
A;Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.
A;Reference number: S70393; MUID:92107687; PMID:1762923
A;Accession: S70395
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: genomic RNA
A;Residues: 1-676 <PER>
A;Cross-references: UNIPROT:P26804; EMBL:Z11128; NID:g61547; PIDN:CAA77479.1; PID:g61549
A;Experimental source: strain FB29
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; polyprotein; spike protein; transmembrane protein

  Query Match            14.9%;  Score 392;  DB 2;  Length 676;
  Best Local Similarity  23.3%;  Pred. No. 4.7e-22;
  Matches  140;  Conservative   81;  Mismatches  221;  Indels  158;  Gaps   22;

Qy          16 PFALTAPPPCCCTTS----------SSPYQEFLXRTRLPGN---IDAPSYRSLSKGNSTF 62
               |::    ||||  :|          : |      |     |   :|  :::| |:|
Db          97 PYSSPPGPPCCSGSSGNVAGCARDCNEPLTSLTPRCNTAWNRLKLDQVTHKS-SEGFYVC 155

Qy          63 TAHTHMPR---NC------YNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFT-HT 112
                  :| ||   :|      | ::  |      ||      ||          | | |
Db         156 PG-SHRPREAKSCGGPDSFYCASWGCETTGRVYW-----KPSS--------SWDYITVDN 201

Qy         113 SMSDGGGIQG----------QAREKQVKEAISQLTRGH---------------------- 140
               :::    :|             |     : ::  | ||
Db         202 NLTSNQAVQVCKDNKWCNPLAIRFTNAGKQVTSWTTGHYWGLRLYVSGQDPGLTFGIRLS 261

Qy         141 -------------------------STPSPYKGLVLSKLHETL-------------RTH 161
                                           | | |    |    ||              |
Db         262 YQNLGPRIPIGPNPVLADQLSFPLPNPLPKPAKSPPASSSTPTLISPSPTPTQPPPAGTG 321

Qy         162 TRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVG 221
                ||::|       |:  :      ||:||      |  : |        |  | ||
Db         322 DRLLNLVQGAYQALNLTNPDKTQECWLCLVSGPPYYEGVAV------LGTYSNHTSA--- 372

Qy         222 PLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC--------------LPSG 267
                  :|  :     ||  :      |    ||  |    : :|               |:|
Db         373 --PANCSVASQHKLTLSEV-----TGRGLCIGTVPKTHQALCNTTLKAGKGSYYLVAPTG 425

Qy         268 IFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI--- 320
                 : | |    ||: :    :   | |  | | :| :    :|:       :|| |:
Db         426 TMWACNTGLTPCLSATVLNRTTDYCVLVELWPRVTYHPPSYVYSQFEKSHRHKREPVSLT 485

Qy         321 LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAA 377
               |  ::    :| :  |:|: ||   :|| : :|   :  |:::|  |:  |:  | ||:
Db         486 LALLLGGLTMGGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLTSLSE 545

Qy         378 VVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTER 437
               ||||||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  ::::
Db         546 VVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLSQRQKLFESSQG 605

Qy         438 WGLLSQWM---PW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
               |     |    ||    :   :||| |:|:|||||||| | ||:||  ||  |:  :: |
Db         606 W--FEGWFNRSPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFVKDRISVVQALVLTQ 663


RESULT 17
VCVWGF
env polyprotein - feline leukemia virus (strain Gardner-Arnstein)
N;Contains: knob protein gp70; R protein; spike protein p15E
```

```
C;Species: feline leukemia virus
A;Note: host Felis sp. (cat)
C;Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 09-Jul-2004
C;Accession: A03991
R;Nunberg, J.H.; Williams, M.E.; Innis, M.A.
J. Virol. 49, 629-632, 1984
A;Title: Nucleotide sequences of the envelope genes to two isolates of feline leukemia v
A;Reference number: A92996; MUID:84115095; PMID:6319767
A;Accession: A03991
A;Molecule type: DNA
A;Residues: 1-662 <NUN>
A;Cross-references: UNIPROT:P03391; GB:K01209; NID:g323910; PIDN:AAA43052.1; PID:g323911
R;Elder, J.H.; McGee, J.S.; Munson, M.; Houghten, R.A.; Kloetzer, W.; Bittle, J.L.; Gran
J. Virol. 61, 8-15, 1987
A;Title: Localization of neutralizing regions of the envelope gene of feline leukemia vi
A;Reference number: A25982; MUID:87061257; PMID:2431166
A;Contents: annotation; peptide synthesis
A;Note: 27 peptides synthesized
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-465/Product: knob protein gp70 #status predicted <KPG>
F;466-645/Product: spike protein p15E #status predicted <SKP>
F;646-662/Product: R protein #status predicted <RRP>
F;43,58,286,322,327,351,354,394,410,430/Binding site: carbohydrate (Asn) (covalent) #sta

  Query Match             14.9%;  Score 391.5;  DB 1;  Length 662;
  Best Local Similarity   27.2%;  Pred. No. 5e-22;
  Matches  116;  Conservative   71;  Mismatches  158;  Indels   81;  Gaps   15;

Qy        124 AREKQVKEAISQLTRGHS-----TP---------SPYKGLVLSKLHETLRTHTRLVSLFN 169
              :|: |::   |::|  ||     ||         :|    |     : : |  ||::|
Db        259 SRQSQIE---SRVTPHHSQGNGGTPGITLVNASIAPLSTPVTPASPKRIGTGDRLINLVQ 315

Qy        170 TTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEI 229
               |   |:        :||:||      |  | :     |:| : |
Db        316 GTYLALNATDPNRTKDCWLCLVSRPPYYEGIAI---LGNYSNQTNPP------------ 359

Qy        230 THTSNLTCVKFSNTIDTTSSQ--CIRWVTPPTRIVC--------------LPSGIFFVCG 273
                 | |:  :   ||   | |  ||  |    : :|               |:| :: |
Db        360 --PSCLSIPQHKLTISEVSGQGLCIGTVPKTHQALCNETQQGHTGAHYLAAPNGTYWACN 417

Qy        274 TSAYHC-----LNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI-------L 321
              |    |     || :|:  | |  | | :| :  : :| |       :| ||       |
Db        418 TGLTPCISMAVLNWTSD-FCVLIELWPRVTYHQPEYVYTHFAKAARFRREPISLTVALML 476

Qy        322 PFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQ 381
                :   |:   :|||  ::  : ||  :|   :: |:: : :|:  |:  | ||: ||||
Db        477 GGLTVGGIAAGVGTGTKALIETAQF-RQLQMAMHTDIQALEESISALEKSLTSLSEVVLQ 535

Qy        382 NRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLL 441
              ||| ||:|  : || |  | || |:| : : :| : : ::|:|:: | :   : : |
Db        536 NRRGLDILFLQEGGLCAALKEECCFYADHTGLVRDNMAKLRERLKQRQQLFDSQQGW--F 593

Qy        442 SQWM---PW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAV-------KLQM 487
                |    ||    :   :|||  |:|:||||||| | ||:||  ||  |       : |
Db        594 EGWFNKSPWFTTLISSIMGPLLILLLILLFGPCILNRLVQFVKDRISVVQALILTQQYQQ 653

Qy        488 VLQMEP 493
              : | :|
Db        654 IKQYDP 659

RESULT 18
A25982
env polyprotein - feline leukemia virus (strain Gardner-Arnstein)
C;Species: feline leukemia virus
C;Date: 25-Oct-1987 #sequence_revision 25-Oct-1987 #text_change 09-Jul-2004
C;Accession: A25982
R;Elder, J.H.; McGee, J.S.; Munson, M.; Houghten, R.A.; Kloetzer, W.; Bittle, J.L.; Gran
J. Virol. 61, 8-15, 1987
```

```
A;Title: Localization of neutralizing regions of the envelope gene of feline leukemia vi
A;Reference number: A25982; MUID:87061257; PMID:2431166
A;Accession: A25982
A;Molecule type: protein
A;Residues: 1-662 <ELD>
A;Cross-references: UNIPROT:Q7LYY7
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei

  Query Match             14.9%;  Score 391.5;  DB 2;  Length 662;
  Best Local Similarity   27.2%;  Pred. No. 5e-22;
  Matches  116;  Conservative   71;  Mismatches  158;  Indels   81;  Gaps   15;

Qy        124 AREKQVKEAISQLTRGHS-----TP---------SPYKGLVLSKLHETLRTHTRLVSLFN 169
              :|: |::   |::|  ||     ||         :|    |     : : |  ||::|
Db        259 SRQSQIE---SRVTPHHSQGNGGTPGITLVNASIAPLSTPVTPASPKRIGTGDRLINLVQ 315

Qy        170 TTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEI 229
               |   |:        :||:||      |  | :     |:| : |
Db        316 GTYLALNATDPNRTKDCWLCLVSRPPYYEGIAI---LGNYSNQTNPP------------ 359

Qy        230 THTSNLTCVKFSNTIDTTSSQ--CIRWVTPPTRIVC--------------LPSGIFFVCG 273
                 | |:  :   ||   | |  ||  |    : :|               |:| :: |
Db        360 --PSCLSIPQHKLTISEVSGQGLCIGTVPKTHQALCNETQQGHTGAHYLAAPNGTYWACN 417

Qy        274 TSAYHC-----LNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI-------L 321
              |    |     || :|:  | |  | | :| :  : :| |       :| ||       |
Db        418 TGLTPCISMAVLNWTSD-FCVLIELWPRVTYHQPEYVYTHFAKAARFRREPISLTVALML 476

Qy        322 PFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQ 381
                :   |:   :|||  ::  : ||  :|   :: |:: : :|:  |:  | ||: ||||
Db        477 GGLTVGGIAAGVGTGTKALIETAQF-RQLQMAMHTDIQALEESISALEKSLTSLSEVVLQ 535

Qy        382 NRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLL 441
              ||| ||:|  : || |  | || |:| : : :| : : ::|:|:: | :   : : |
Db        536 NRRGLDILFLQEGGLCAALKEECCFYADHTGLVRDNMAKLRERLKQRQQLFDSQQGW--F 593

Qy        442 SQWM---PW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAV-------KLQM 487
                |    ||    :   :|||  |:|:||||||| | ||:||  ||  |       : |
Db        594 EGWFNKSPWFTTLISSIMGPLLILLLILLFGPCILNRLVQFVKDRISVVQALILTQQYQQ 653

Qy        488 VLQMEP 493
              : | :|
Db        654 IKQYDP 659


RESULT 19
VCMVRV
env polyprotein precursor - Rauscher mink cell focus-forming virus
N;Contains: coat protein p12E; coat protein p15E; knob protein gp70
C;Species: Rauscher mink cell focus-forming virus
C;Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 09-Jul-2004
C;Accession: A03990
R;Vogt, M.; Haggblom, C.; Swift, S.; Haas, M.
J. Virol. 55, 184-192, 1985
A;Title: Envelope gene and long terminal repeat determine the different biological prope
A;Reference number: A93011; MUID:85237696; PMID:4009793
A;Accession: A03990
A;Molecule type: DNA
A;Residues: 1-640 <VOG>
A;Cross-references: UNIPROT:P06445; GB:M10100; NID:g332068; PIDN:AAA46528.1; PID:g332070
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-443/Product: knob protein gp70 #status predicted <KBP>
F;444-640/Product: coat protein p15E #status predicted <PFE>
F;444-623/Product: coat protein p12E #status predicted <PTE>
F;43,58,300,332,339,372,408,576/Binding site: carbohydrate (Asn) (covalent) #status pred

  Query Match             14.8%;  Score 390.5;  DB 1;  Length 640;
  Best Local Similarity   29.1%;  Pred. No. 5.7e-22;
  Matches  105;  Conservative   63;  Mismatches  146;  Indels   47;  Gaps   10;

Qy         160 THTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVL 219
               |  ||::| :     |:  |      ||:||      |  : |        |  | ||
Db         284 TGDRLLNLVDGAYQALNLTSPDKTQECWLCLVAEPPYYEGVAV------LGTYSNHTSA- 336

Qy         220 VGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVC--------------LP 265
                    :|  :     ||  :      |    ||  |    : :|               |
Db         337 ----PTNCSVASQHKLTLSEV-----TGQGLCIGTVPKTHQALCNTTLKTNKGSYYLVAP 387

Qy         266 SGIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI- 320
               :|  : | |    ||: :    :   | |  | | :| :    :|:       :|| |:
Db         388 AGTTWACNTGLTPCLSATVLNRTTDYCVLVELWPRVTYHPPSYVYSQFEKSYRHKREPVS 447

Qy         321 --LPFVIRAGVLGRLGTGIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSL 375
                 |  ::    :| :  |:|: ||   :|| : :|   :  |:::|  |:  |:  | ||
Db         448 LTLALLLGGLTMGGIAAGVGTGTTALVATQQFQQLHAAVQDDLKEVEKSITNLEKSLTSL 507

Qy         376 AAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNT 435
               : ||||||| ||||  | || |  | || |:| : : :| : : ::|:|:  | :  :::
Db         508 SEVVLQNRRGLDLLFLKEGGLCAALKEECCFYADHTGLVRDSMAKLRERLTQRQKLFESS 567

Qy         436 ERW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVL 489
               : |  || ::  ||    :   :|||  |:|:||||||| | ||:||  ||  |:  ::
Db         568 QGWFEGLFNR-SPWFTTLISTIMGPLIILLLILLFGPCILNRLVQFVKDRISVVQALVLT 626

Qy         490 Q 490
               |
Db         627 Q 627

RESULT 20
VCVWEK
env polyprotein - AKV murine leukemia virus
N;Contains: knob protein gp76; R protein; spike protein p15E
C;Species: AKV murine leukemia virus
C;Date: 05-Apr-1983 #sequence_revision 03-Aug-1984 #text_change 09-Jul-2004
C;Accession: A92995; A93448; A03984
R;Herr, W.
J. Virol. 49, 471-478, 1984
A;Title: Nucleotide sequence of AKV murine leukemia virus.
A;Reference number: A92995; MUID:84115072; PMID:6319746
A;Accession: A92995
A;Molecule type: genomic RNA
A;Residues: 1-669 <HER>
A;Cross-references: UNIPROT:P03386; GB:J01998; GB:J01999; GB:K00016; GB:K00017; GB:K0001
R;Herr, W.; Corbin, V.; Gilbert, W.
Nucleic Acids Res. 10, 6931-6944, 1982
A;Title: Nucleotide sequence of the 3' half of AKV.
A;Reference number: A93448; MUID:83090450; PMID:6294621
A;Accession: A93448
A;Molecule type: DNA
A;Residues: 1-34,'R',36-462,'K',464-591,'K',593-669 <HE2>
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-470/Product: knob protein gp76 #status predicted <KNB>
F;471-650/Product: spike protein p15E #status predicted <SPK>
F;651-669/Product: R protein #status predicted <RPT>
F;43,199,327,359,399/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             14.8%;  Score 390;  DB 1;  Length 669;
  Best Local Similarity   24.1%;  Pred. No. 6.6e-22;
  Matches  141;  Conservative   82;  Mismatches  229;  Indels  132;  Gaps   19;

Qy          16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGNIDAPSYRSLSKGNSTFTAH 65
               ||:    ||||  :| |          |   :  |     |        |||       |
Db          94 PFSPPPGPPCCSGSSDSTPGCSRDCEEPLTSYTPRCNTAWN-----RLKLSK-----VTH 143

Qy          66 THM---------------------PRNCYNSATLCMHANTHYW---------------TGK 90
                |                     | : | ::  |       |               |
```

```
Db         144 AHNGGFYVCPGPHRPRWARSCGGPESFYCASWGCETTGRASWKPSSSWDYITVSNNLTSD 203

Qy          91 MINPSCPG--------------GLGATV-----CWTYFTHTSMSDGG---GIQGQAREKQ 128
                  | | |              |  ||       |    : |  | |   ||: :  :
Db         204 QATPVCKGNEWCNSLTIRFTSFGKQATSWVTGHWWGLRLYVSGHDPGLIFGIRLKITDSG 263

Qy         129 VKEAIS-----QLTRGHSTPSPYKGLVLSKLHET-----------LRTHTRLVSLFNTTL 172
                :  |         |  | | | :    |    |                ||::|
Db         264 PRVPIGPNPVLSDRRPPSRPRPTRSPPPSNSTPTETPLTLPEPPPAGVENRLLNLVKGAY 323

Qy         173 TRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT-- 230
                 |:  |      ||:||      |  : |   ::| ::     ||     ::  |:|
Db         324 QALNLTSPDKTQECWLCLVSGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQ 383

Qy         231 --------HTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNG 282
                        |  : |    ||   ||          :  :  |:|  : |  |    |::
Db         384 GLCIGAVPKTHQVLC----NTTQKTSD--------GSYYLAAPTGTTWACSTGLTPCIST 431

Qy         283 S----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGT 335
                :   :   | |  | | :| ::   :|:    :   || |:   |  ::    :| :
Db         432 TILDLTTDYCVLVELWPRVTYHSPSYVYHQFERRAKYKREPVSLTLALLLGGLTMGGIAA 491

Qy         336 GIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAK 392
               |:|: ||   :|| : :|   :: |:::|  |:  |:  | ||: ||||||| ||||  |
Db         492 GVGTGTTALVATQQFQQLQAAMHDDLKEVEKSITNLEKSLTSLSEVVLQNRRGLDLLFLK 551

Qy         393 RGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW--- 447
                || |  | || |:| : : :| : : ::|:|:  | :  :: : |  || ::  ||
Db         552 EGGLCAALKEECCFYADHTGLVRDSMAKLRERLSQRQKLFESQQGWFEGLFNK-SPWFTT 610

Qy         448 -VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
                :   :|||  |:|:||||||| | ||:|:  ||  |:  :: |
Db         611 LISTIMGPLIILLLILLFGPCILNRLVQFIKDRISVVQALVLTQ 654

RESULT 21
VCMVLB
env polyprotein - feline leukemia virus (strain lambda-B1) (fragment)
N;Contains: coat protein gp70; coat protein p15E
C;Species: feline leukemia virus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 09-Jul-2004
C;Accession: A27172
R;Nicolaisen-Strouss, K.; Kumar, H.P.M.; Fitting, T.; Grant, C.K.; Elder, J.H.
J. Virol. 61, 3410-3415, 1987
A;Title: Natural feline leukemia virus variant escapes neutralization by a monoclonal ant
A;Reference number: A27172; MUID:88036192; PMID:2444714
A;Accession: A27172
A;Molecule type: DNA
A;Residues: 1-662 <NIC>
A;Cross-references: UNIPROT:P11261; GB:J03448; NID:g323896; PIDN:AAA43048.1; PID:g323897
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-465/Product: coat protein gp70 #status predicted <GPS>
F;466-662/Product: coat protein p15E #status predicted <GPT>
F;43,58,286,322,327,351,354,430/Binding site: carbohydrate (Asn) (covalent) #status pred

  Query Match             14.8%;  Score 389.5;  DB 1;  Length 662;
  Best Local Similarity   27.8%;  Pred. No. 7.1e-22;
  Matches  115;  Conservative   73;  Mismatches  169;  Indels   57;  Gaps   15;

Qy         124 AREKQVKEAISQLTRGHS-----TP---------SPYKGLVLSKLHETLRTHTRLVSLFN 169
               :|: |::   |::|  ||     ||         :|    |     : : |  ||::|
Db         259 SRQSQIE---SRVTPHHSQGNGGTPGITLVNASIAPLSTPVTPASPKRIGTGNRLINLVQ 315

Qy         170 TTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEI 229
                |   |:  :     :||:||      |  | |     |:| : |     :      | |
Db         316 GTYLALNVTNPNKTKDCWLCLVSRPPYYEGIAV---LGNYSNQTNPPPSCLSDPQHKLTI 372

Qy         230 THTSNL-TCVKFSNTIDTTSSQCIRWVT---PPTRIVCLPSGIFFVCGTSAYHC-----L 280
               :  |   :|:    |:  |     :        |  :  ||| :: | |    |     |
Db          373 SEVSGQGSCI---GTVPKTHQALCKKTQKGHKGTHYLAAPSGTYWACNTGLTPCISMAVL 429

Qy          281 NGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI-------LPFVIRAGVLGRL 333
                | :|:  | |  | | :| :  : :| |       :| ||       |  :   |:   :
Db          430 NWTSD-FCVLIELWPRVTYHQPEYVYTHFDKTVRLRREPISLTVALMLGGLTVGGIAAGV 488

Qy          334 GTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKR 393
                |||  ::  : ||  :|   :: |:: : :|:  |:  | ||: ||||||| ||:|  :
Db          489 GTGTKALLETAQF-GQLQMAMHTDIQALEESISALEKSLTSLSEVVLQNRRGLDILFLQE 547

Qy          394 GGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWM---PW--- 447
                || |  | || |:| : : :| : : ::|:|:: | :   : : |     |    ||
Db          548 GGLCAALKEECCFYADHTGLVRDNMAKLRERLKQRQQLFDSQQGW--FEGWFNKSPWFTT 605

Qy          448 -VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAV-------KLQMVLQMEP 493
                 :   :|||  |:|:||||||| | ||:||  ||  |       : | : | :|
Db          606 LISSIMGPLLILLLILLFGPCILNRLVQFVKDRISVVQALILTQQYQQIKQYDP 659

RESULT 22
VCVWM2
env polyprotein - Friend mink cell focus-forming virus
N;Contains: knob protein gp70; spike protein p15E
C;Species: Friend mink cell focus-forming virus
C;Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 16-Jul-1999
C;Accession: A03989
R;Koch, W.; Zimmermann, W.; Oliff, A.; Friedrich, R.
J. Virol. 49, 828-840, 1984
A;Title: Molecular analysis of the envelope gene and long terminal repeat of Friend mink
A;Reference number: A03989; MUID:84138778; PMID:6321768
A;Accession: A03989
A;Molecule type: DNA
A;Residues: 1-627 <KOC>
A;Cross-references: GB:M12528; NID:g331918; PIDN:AAA46483.1; PID:g331920
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-431/Product: knob protein gp70 #status predicted <KNB>
F;432-627/Product: spike protein p15E #status predicted <SPK>
F;43,58,288,320,360,364,396/Binding site: carbohydrate (Asn) (covalent) #status predicte

  Query Match             14.7%;  Score 386;  DB 1;  Length 627;
  Best Local Similarity   25.2%;  Pred. No. 1.2e-21;
  Matches  129;  Conservative   79;  Mismatches  184;  Indels  120;  Gaps   17;

Qy           54 SLSKGNSTFTAHTHMPRN---CYNSATLCMHANTHYWTGKMINPSCPGG----------- 99
                || :||:        |||   ||:|: :             |  : |||
Db          149 SLKRGNT--------PRNQGPCYDSSAV----------SSGIQGATPGGRCNPLVLEFTD 190

Qy          100 LGATVCWTYFTHTSMSDGGGIQGQ--------------AREKQVKEAISQLTRGHS---- 141
                 |    |         ||  : |               :  :||     ::  | :
Db          191 AGKKASW---------DGPKVWGLRLYRSTGTDPVTRFSLTRQVLNIGPRIPIGPNPVIT 241

Qy          142 -----------------TPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPT 184
                                  |||     :        |  ||::| :     |:  |
Db          242 DQLPPSRPVQIMLPRPPQPSPTGAASIQP-----GTGDRLLNLVDGAYQALNLTSPDKTQ 296

Qy          185 NCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT----------HTSN 234
                 ||:||      |  : |   ::| ::     ||     ::  |:|           |
Db          297 ECWLCLVSGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQGLCVGAVPKTHQ 356

Qy          235 LTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGS----SESMCFL 290
                  |    ||   ||          :  : |:|  : | |    ||: :    :   | |
Db          357 ALC----NTTQNTSD--------GSYYLAAPAGTIWACNTGLTPCLSTTVLNLTTDYCVL 404

Qy          291 SFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGTGIGSITT---ST 344
                  | | :| ::   :|     |   ||  |:   |  ::    :| :  |:|: ||   :|
Db          405 VELWPKVTYHSPGYVYGQFERKTKYKREPVSLTLALLLGGLTMGGIAAGVGTGTTALVAT 464

Qy          345 QFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEER 404
                | : :|   :  |:::|  |:  |:  | ||: ||||||| ||||  |  | |  | ||

Db          465 QQFQQLHAAVQDDLKEVEKSITNLEKSLTSLSEVVLQNRRGLDLLFLKERGLCAALKEEC 524

Qy          405 CYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW----VLPFLGPLAAL 458
                |:| : : :| : : ::|:|:  | :  :::: |  || ::  ||    :   :|||  |
Db          525 CFYADHTGLVRDSMAKLRERLTQRQKLFESSQGWFEGLFNR-SPWFTTLISTIMGPLIIL 583

Qy          459 ILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
                :|:||||||| | ||:||  ||  |:  :: |
Db          584 LLILLFGPCILNRLVQFVKDRISVVQALVLTQ 615

RESULT 23
VCMVFP
env polyprotein precursor (clone CFE-6) - feline leukemia virus (provirus)
N;Contains: coat protein gp70; coat protein p15E
C;Species: feline leukemia virus
A;Note: host Felis silvestris catus (domestic cat)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C;Accession: A31479
R;Kumar, D.V.; Berry, B.T.; Roy-Burman, P.
J. Virol. 63, 2379-2384, 1989
A;Title: Nucleotide sequence and distinctive characteristics of the env gene of endogenou
A;Reference number: A31479; MUID:89199802; PMID:2539525
A;Accession: A31479
A;Molecule type: DNA
A;Residues: 1-668 <KUM>
A;Cross-references: GB:M25425; NID:g163849; PIDN:AAA30809.1; PID:g163850
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-465/Product: coat protein gp70 #status predicted <GUP>
F;466-668/Product: coat protein p15E #status predicted <TPP>
F;43,58,286,322,327,351,354,394,410,430,597/Binding site: carbohydrate (Asn) (covalent)

  Query Match             14.6%;  Score 385;  DB 1;  Length 668;
  Best Local Similarity   28.6%;  Pred. No. 1.6e-21;
  Matches  120;  Conservative   72;  Mismatches  147;  Indels   80;  Gaps   16;

Qy          111 HTSMSDGG--GIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLF 168
                |    :||  ||        |  :|: |:   :  ||          : : |  ||::|
Db          271 HHPQGNGGTPGI------TLVNASIAPLSTPVTPASP----------KRIGTGNRLINLV 314

Qy          169 NTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLE 228
                  |   |:  :     :||:||      |  | |     |:| : |
Db          315 QGTYLTLNVTNPNKTKDCWLCLVSRPPYYEGIAV---LGNYSNQTNPP------------ 359

Qy          229 ITHTSNLTCVKFSNTIDTTSSQ--CIRWVTPPTRIVC--------------LPSGIFFVC 272
                    | |:  :   ||   | |  ||  |    : :|               |:| :: |
Db          360 ---PSCLSVPQHKLTISEVSGQGLCIATVPKTHQALCNKTQKGHRGTHYLVAPNGTYWAC 416

Qy          273 GTSAYHC-----LNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIRA 327
                 |    |     || :|:  | |: | | :| :  : :|:|   ||  || ||   |  |
Db          417 NTGLTPCISMAVLNWTSD-FCVLTELWPRITYHEPEYIYSHFENKPRFKRDPISLTV--A 473

Qy          328 GVLGRLGTGIGSITTS---------TQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAV 378
                 :||  |  :| :  :         |  :  :| ::   |:: : :|:  |:  | ||: |
Db          474 LMLG--GITVGGMARNRNRDCGLLETAQFRQLQMAMHTDIQALEESISALEKSLTSLSEV 531

Qy          379 VLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW 438
                |||||| ||:|  : || |  | || |:| : : :| : : ::|:|:: | ::| ::::
Db          532 VLQNRRGLDILFLQEGGLCTALKEECCFYADHTGLVRDNMAKLRERLKQR-QQLFDSQQD 590

Qy          439 GLLSQWM---PW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
                | |  |    ||    :   :|||  |:|:||||||| | ||:||  ||  |:  ::  |
Db          591 G-LEGWFNKSPWFTTLISSIMGPLMILLLILLFGPCILNRLVQFVKDRISVVQTLVLTQ 648

RESULT 24
A46511
envelope protein - AKV murine leukemia virus
C;Species: AKV murine leukemia virus
```

```
C;Date: 18-Jun-1993 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: A46511
R;Hayashi, H.; Matsubara, H.; Yokota, T.; Kuwabara, I.; Kanno, M.; Koseki, H.; Isono, K.
J. Immunol. 149, 1223-1229, 1992
A;Title: Molecular cloning and characterization of the gene encoding mouse melanoma anti
A;Reference number: A46511; MUID:92364323; PMID:1380036
A;Accession: A46511
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-669 <HAY>
A;Cross-references: UNIPROT:Q9Q9A5; UNIPROT:Q9DQ24; UNIPROT:Q83382
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:110845, NCBIP:110846)
C;Superfamily: type C retrovirus env polyprotein

  Query Match             14.6%;  Score 385;  DB 2;  Length 669;
  Best Local Similarity   24.3%;  Pred. No. 1.6e-21;
  Matches  142;  Conservative   79;  Mismatches  231;  Indels  132;  Gaps   19;

Qy         16 PFALTAPPPCCCTTSSS----------PYQEFLXRTRLPGNIDAPSYRSLSKGNSTFTAH 65
              ||:    ||||  :| |          |   :  |     |        |||       |
Db         94 PFSPPPGPPCCSGSSDSTSGCSRDCEEPLTSYTPRCNTAWN-----RLKLSK-----VTH 143

Qy         66 THM--------------------PRNCYNSATLCMHANTHYW---------------TGK 90
               |                     | : | ::  |       |               |
Db        144 AHNEGFYVCPGPHRPRWARSCGGPESFYCASWGCETTGRASWKPSSSWDYITVSNNLTSD 203

Qy         91 MINPSCPG--------------GLGATV-----CWTYFTHTSMSDGG---GIQGQAREKQ 128
                 | | |              |  ||       |    : |  | |   ||: :  :
Db        204 QATPVCKGNKWCNSLTIRFTSFGKQATSWVTGHWWGLRLYVSGHDPGLIFGIRLKITDSG 263

Qy        129 VKEAIS-----QLTRGHSTPSPYKGLVLSKLHET-----------LRTHTRLVSLFNTTL 172
               :  |         |  | | | :    |    |                ||::|
Db        264 PRVPIGPNPVLSDRRPPSRPRPTRSPPPSNSTPTETPLTLPEPPPAGVENRLLNLVKGAY 323

Qy        173 TRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT-- 230
                |:  |      ||:||      |  : |   ::| ::     ||     ::  |:|
Db        324 QALNLTSPDKTQECWLCLVSGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSEVTGQ 383

Qy        231 --------HTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNG 282
                       |  : |    ||   ||      |         |:|  : | |    |::
Db        384 GLCIGAVPKTHQVLC----NTTQKTSDGSYYLVA--------PTGTTWACSTGLTPCIST 431

Qy        283 S----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI---LPFVIRAGVLGRLGT 335
              :    :   | |  | | :| ::   :|:    :   || |:   |  ::    :| :
Db        432 TILNLTTDYCVLVELWPRVTYHSPSYVYHQFERRAKYKREPVSLTLALLLGGLTMGGIAA 491

Qy        336 GIGSITT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAK 392
              |:|: ||   :|| : :|   :: |:::|  |:  |:  | ||: ||||||| ||||  |
Db        492 GVGTGTTALVATQQFQQLQAAMHDDLKEVEKSITNLEKSLTSLSEVVLQNRRGLDLLFLK 551

Qy        393 RGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW--- 447
               || |  | || |:| : : :| : : ::| |:  | :  :: : |  ||  ::  ||
Db        552 EGGLCAALKEECCFYADHTGLVRDSMAKLRFRLSQRQKLFESQQGWFEGLFNK-SPWFTT 610

Qy        448 -VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
               :   :|||  |:|:||||||| | ||:|:  ||  |:  :: |
Db        611 LISTIMGPLIILLLILLFGPCILNRLVQFIKDRISVVQALVLTQ 654

RESULT 25
VCVWM1
env polyprotein - mink cell focus-forming virus
N;Contains: knob protein gp70; R protein; spike protein p15E
C;Species: mink cell focus-forming virus
C;Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 16-Jul-1999
C;Accession: A03987; B03987
R;Mark, G.E.; Rapp, U.R.
J. Virol. 49, 530-539, 1984
A;Title: Envelope gene sequence of two in vitro-generated mink cell focus-forming murine
A;Reference number: A03987; MUID:84115078; PMID:6319752
A;Accession: A03987
A;Molecule type: DNA
```

```
A;Residues: 1-640 <MAR>
A;Cross-references: GB:K02725; NID:g331614; PIDN:AAA46375.1; PID:g331616
A;Experimental source: clone CI-3
A;Accession: B03987
A;Molecule type: DNA
A;Residues: 1-314,543-640 <MA2>
A;Experimental source: clone CI-4
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-441/Product: knob protein gp70 #status predicted <KNB>
F;442-621/Product: spike protein #status predicted <SPK>
F;622-640/Product: R protein #status predicted <RPT>
F;43,58,297,329,336,369/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             14.6%;  Score 384;  DB 1;  Length 640;
  Best Local Similarity   24.7%;  Pred. No. 1.8e-21;
  Matches  128;  Conservative   80;  Mismatches  187;  Indels  124;  Gaps   18;

Qy         54 SLSKGNSTFTAHTHMPRN---CYNSATLCMHANTHYWTGKMINPSCPGG----------- 99
              || :||:        |||   ||:|: :             |  : |||
Db        149 SLKRGNT--------PRNQGPCYDSSAV----------SSDIKGATPGGRCNPLVLEFTD 190

Qy        100 LGATVCWTYFTHTSMSDGGGIQG-----------QAREKQVKEAIS-------------- 134
               |    |         ||  : |             |    :: ::
Db        191 AGKKASW---------DGPKVWGLRLYRSTGTDPVTRFSLTRQVLNIGPRVPIGPNPVIT 241

Qy        135 -----------QLTRGHSTPSPYKGLVLSKL---HETLRTHTRLVSLFNTTLTRLHEVSA 180
                          | |    | |    :: :     : | |  ||::| |     |:  |
Db        242 DQLPPSRPVQIMLPRPPQPPPPGAASIVPETAPPSQQLGTGDRLLNLVNGAYQALNLTSP 301

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLV------------GPLVSNLE 228
                   ||:||      |  : |   ::| ::     ||              |  ::
Db        302 DKTQECWLCLVAGPPYYEGVAVLGTYSNHTSAPANCSVASQHKLTLSGVAGRGLCIAAFP 361

Qy        229 ITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGS----S 284
               || :        ||   ||             :  |:|  : | |    ||: :    :
Db        362 KTHQA------LCNTTQKTSDGSYH--------LAAPAGTIWACNTGLTPCLSTTVLDLT 407

Qy        285 ESMCFLSFLVPPMTIYTEQDLYNHVVPKPHN-KRVPI---LPFVIRAGVLGRLGTGIGSI 340
                 | |  | | :| ::   :|     |    || |:   |  ::    :| :  |:|:
Db        408 TDYCVLVELWPKVTYHSPSYVYGQFEKKKTKYKREPVSLTLALLLGGLTMGGIAAGVGTG 467

Qy        341 TT---STQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTC 397
              ||   :|| : :|   :: |:::|  |:  |:  | ||: ||||||| ||||  | || |
Db        468 TTALVATQQFQQLQAAMHDDLKEVEKSITNLEKSLTSLSEVVLQNRRGLDLLFLKEGGLC 527

Qy        398 LFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERW--GLLSQWMPW----VLPF 451
                | || |:| : : :| : : ::|:|:  | :  :: : |  || ::  ||    :
Db        528 AALKEECCFYADHTGLVRDSMAKLRERLSQRQKLFESQQGWFEGLFNK-SPWFTTLISTI 586

Qy        452 LGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              :|||  |:|:||||| | | ||:|:  ||  |:  :: |
Db        587 MGPLIILLLILLFGPWILNRLVQFIKDRISVVQALVLTQ 625

RESULT 26
VCMVS2
env polyprotein precursor - feline sarcoma virus (strain GA)
N;Contains: coat protein gp70; coat protein p15E
C;Species: feline sarcoma virus
A;Note: host Felis silvestris catus (domestic cat)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 12-Apr-1996
C;Accession: B33741
R;Guilhot, S.; Hampe, A.; D'Auriol, L.; Galibert, F.
Virology 161, 252-258, 1987
A;Title: Nucleotide sequence analysis of the LTRs and env genes of SM-FeSV and GA-FeSV.
A;Reference number: A33741; MUID:88044502; PMID:2823466
A;Accession: B33741
A;Molecule type: DNA
A;Residues: 1-642 <GUI>
C;Genetics:
A;Gene: env
```

```
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-445/Product: coat protein gp70 #status predicted <GUP>
F;446-642/Product: coat protein p15E #status predicted <GUQ>
F;34,42,57,267,302,307,331,334,374,390,410,578/Binding site: carbohydrate (Asn) (covalen

  Query Match             14.5%;  Score 383;  DB 1;  Length 642;
  Best Local Similarity   28.2%;  Pred. No. 2.1e-21;
  Matches  106;  Conservative   62;  Mismatches  144;  Indels   64;  Gaps   12;

Qy        160 THTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVL 219
              |  ||::|   |   |:        :||:||      |  | :     |:| : |
Db        286 TGDRLINLVQGTYLALNATDPNRTKDCWLCLVSRPPYYEGIAI---LGNYSNQTNPP--- 339

Qy        220 VGPLVSNLEITHTSNLTCVKFSNTIDTTSSQ--CIRWVTPPTRIVC-------------- 263
                           | |:  :   ||   | |  ||  |    : :|
Db        340 ------------PSCLSIPQHKLTISEVSGQGLCIGTVPKTHQALCNETQQGHTGAHYLA 387

Qy        264 LPSGIFFVCGTSAYHC-----LNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRV 318
               |:| :: | |    |     || :|:  | |  | | :| :  : :| |       :|
Db        388 APNGTYWACNTGLTPCISMAVLNWTSD-FCVLIELWPRVTYHQPEYVYTHFAKAARFRRE 446

Qy        319 PI-------LPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQ 371
              ||       |  :   |:   :|||  ::  : ||  :|   :: |:: : :|:  |:
Db        447 PISLTVALMLGGLTVGGIAAGVGTGTKALIETAQF-RQLQMAMHTDIQALEESISALEKS 505

Qy        372 LNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEE 431
              | ||: ||||||| ||:|  : || |  | || |:| : : :| : : ::|:|:: | :
Db        506 LTSLSEVVLQNRRGLDILFLQEGGLCAALKEECCFYADHTGLVRDNMAKLRERLKQRQQL 565

Qy        432 LQNTERWGLLSQWM---PW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAV- 483
                : : |     |    ||    :   :|||  |:|:||||||| | ||:||  ||  |
Db        566 FDSQQGW--FEGWFNKSPWFTTLISSIMGPLLILLLILLFGPCILNRLVQFVKDRISVVQ 623

Qy        484 ------KLQMVLQMEP 493
                    : | : | :|
Db        624 ALILTQQYQQIKQYDP 639

RESULT 27
VCMVSS
env polyprotein precursor - feline sarcoma virus (strain SM)
N;Contains: coat protein gp70; coat protein p15E
C;Species: feline sarcoma virus
A;Note: host Felis silvestris catus (domestic cat)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 12-Apr-1996
C;Accession: A33741
R;Guilhot, S.; Hampe, A.; D'Auriol, L.; Galibert, F.
Virology 161, 252-258, 1987
A;Title: Nucleotide sequence analysis of the LTRs and env genes of SM-FeSV and GA-FeSV.
A;Reference number: A33741; MUID:88044502; PMID:2823466
A;Accession: A33741
A;Molecule type: DNA
A;Residues: 1-645 <GUI>
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-36/Domain: signal sequence #status predicted <SIG>
F;37-448/Product: coat protein gp70 #status predicted <GUP>
F;449-645/Product: coat protein p15E #status predicted <GUQ>
F;38,46,61,270,305,310,334,337,377,393,413,581/Binding site: carbohydrate (Asn) (covalen

  Query Match             14.5%;  Score 383;  DB 1;  Length 645;
  Best Local Similarity   27.5%;  Pred. No. 2.2e-21;
  Matches  114;  Conservative   70;  Mismatches  161;  Indels   70;  Gaps   15;

Qy        124 AREKQVKEAIS---QLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              :|: | |  ::     | | :||       :|   : : |  ||::|   |   |:
Db        253 SRQSQTKSKVATQKPQTNG-TTPRSVAPATMSP--KRIGTRDRLINLVQGTYLALNATDP 309

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
                  :||:||      |  | :     |:| : |                  | |:  :
Db        310 NKTKDCWLCLVSRPPYYEGIAI---LGNYSNQTNPP---------------PSCLSTPQH 351

Qy        241 SNTIDTTSSQ--CIRWVTPPTRIVC--------------LPSGIFFVCGTSAYHC----- 279
                ||   | |  ||  |    : :|               |:| :: | |    |
Db        352 KLTISEVSGQGLCIGTVPRTHQALCNKTQQGHTGAHYLAAPNGTYWACNTGLTPCISMAV 411

Qy        280 LNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI-------LPFVIRAGVLGR 332
              || :|:  | |  | | :| :  : :| |       :| ||       |  :   |:
Db        412 LNWTSD-FCVLIELWPRVTYHQPEYIYTHFDKAVRFRREPISLTVALMLGGLTVGGIAAG 470

Qy        333 LGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAK 392
              :|||  ::  : ||  :|   :: |:: : :|:  |:  | ||: ||||||| ||:|  :
Db        471 VGTGTKALLETAQF-RQLQIAMHTDIQALEESISALEKSLTSLSEVVLQNRRGLDILFLQ 529

Qy        393 RGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWM---PW-- 447
               || |  | || |:| : : :| : : ::|:|:: | :   : : |     |    ||
Db        530 GGGLCAALKEECCFYADHTGLVRDNMAKLRERLKQRQQLFDSQQGW--FEGWFNKSPWFT 587

Qy        448 --VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAV-------KLQMVLQMEP 493
                :   :|||  |:|:||||||| | ||:||  ||  |       : | : | :|
Db        588 TLISSIMGPLLILLLILLFGPCILNRLVQFVKDRISVVQALILTQQYQQIQQYDP 642

RESULT 28
T10533
env polyprotein precursor - feline leukemia virus (strain FeLV-FAIDS)
N;Contains: env protein gp70; env protein p15E
C;Species: feline leukemia virus
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T10533
R;Donahue, P.R.; Hoover, E.A.; Beltz, G.A.; Riedel, N.; Hirsch, V.M.; Overbaugh, J.; Mull
J. Virol. 62, 722-731, 1988
A;Title: Strong sequence conservation among horizontally transmissible, minimally pathoge
A;Reference number: Z17078; MUID:88119207; PMID:2828667
A;Accession: T10533
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-642 <DON>
A;Cross-references: UNIPROT:Q85522; EMBL:M18247; NID:g323904; PIDN:AAA93093.1; PID:g3239
C;Superfamily: type C retrovirus env polyprotein

  Query Match             14.5%;  Score 382;  DB 2;  Length 642;
  Best Local Similarity   28.2%;  Pred. No. 2.6e-21;
  Matches  106;  Conservative   62;  Mismatches  144;  Indels   64;  Gaps   12;

Qy        160 THTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVL 219
              |  ||::|   |   |:        :||:||      |  | :     |:| : |
Db        286 TGDRLINLVQGTYLALNATDPNKTKDCWLCLVSRPPYYEGIAI---LGNYSNQTNPP--- 339

Qy        220 VGPLVSNLEITHTSNLTCVKFSNTIDTTSSQ--CIRWVTPPTRIVC-------------- 263
                           | |:  :   ||   | |  ||  |    : :|
Db        340 ------------PSCLSIPQHKLTISEVSGQGLCIGTVPKTHQALCNKTQQGHTGAHYLA 387

Qy        264 LPSGIFFVCGTSAYHC-----LNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRV 318
               |:| :: | |    |     || :|:  | |  | | :| :  : :| |       :|
Db        388 APNGTYWACNTGLTPCISMAVLNWTSD-FCVLIELWPRVTYHQPEYVYTHFAKAVRFRRE 446

Qy        319 PI-------LPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQ 371
              ||       |  :   |:   :|||  ::  : ||  :|   :: |:: : :|:  |:
Db        447 PISLTVALMLGGLTVGGIAAGVGTGTKALLETAQF-RQLQMAMHTDIQALEESISALEKS 505

Qy        372 LNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEE 431
              | ||: ||||||| ||:|  : || |  | || |:| : : :| : : ::|:|:: | :
Db        506 LTSLSEVVLQNRRGLDILFLQEGGLCAALKEECCFYADHTGLVRDNMAKLRERLKQRQQL 565

Qy        432 LQNTERWGLLSQWM---PW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAV- 483
                : : |     |    ||    :   :|||  |:|:||||||| | ||:||  ||  |
Db        566 FDSQQGW--FEGWFNRSPWFTTLISSIMGPLLILLLILLFGPCILNRLVQFVKDRISVVQ 623

Qy        484 ------KLQMVLQMEP 493
                    : | : | :|
Db        624 ALILTQQYQQIKQYDP 639
```

```
RESULT 29
VCMVSA
env polyprotein precursor - feline leukemia virus (strain Sarma)
N;Alternate names: coat polyprotein
N;Contains: knob protein gp70; spike protein p15E
C;Species: feline leukemia virus
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 09-Jul-2004
C;Accession: A29013
R;Riedel, N.; Hoover, E.A.; Gasper, P.W.; Nicolson, M.O.; Mullins, J.I.
J. Virol. 60, 242-250, 1986
A;Title: Molecular analysis and pathogenesis of the feline aplastic anemia retrovirus, f
A;Reference number: A29013; MUID:86308240; PMID:3018287
A;Accession: A29013
A;Molecule type: DNA
A;Residues: 1-639 <RIE>
A;Cross-references: UNIPROT:P06752; GB:M14331; NID:g323898; PIDN:AAA43049.1; PID:g323899
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-442/Product: knob protein #status predicted <NOB>
F;443-639/Product: spike protein #status predicted <SPK>
F;35,43,58,299,304,328,331,387,407,575/Binding site: carbohydrate (Asn) (covalent) #stat

  Query Match             14.4%;  Score 380.5;  DB 1;  Length 639;
  Best Local Similarity   27.6%;  Pred. No. 3.3e-21;
  Matches  112;  Conservative   69;  Mismatches  160;  Indels   65;  Gaps   14;

Qy          124 AREKQVKEAIS----QLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVS 179
                :|: | |  ::    |:|   |||       :    : : |  ||::|   |   |:
Db          247 SRQSQTKSKVTTQRPQIT--SSTPRSVASATMGP--KRIGTGDRLINLVQGTYLALNATD 302

Qy          180 AQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVK 239
                     :||:||      |  | |     |:| : |                  | |:  :
Db          303 PNKTKDCWLCLVSRPPYYEGIAV---LGNYSNQTNPP---------------PSCLSTPQ 344

Qy          240 FSNTIDTTSSQ--CIRWVTPPTRIVC--------------LPSGIFFVCGTSAYHC---- 279
                   ||   | |  ||  |    : :|               |:| :: | |    |
Db          345 HKLTISEVSGQGLCIGTVPKTHQALCKKTQKGHKGTHYLAAPNGTYWACNTGLTPCISMA 404

Qy          280 -LNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI-------LPFVIRAGVLG 331
                 || :|:  | |  | | :| :  : :| |       :| ||       |  :   |:
Db          405 VLNWTSD-FCVLIELWPRVTYHQPEYIYTHFDKAVRFRREPISLTVALMLGGLTVGGIAA 463

Qy          332 RLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTA 391
                 :|||  ::  : ||  :|   :: |:: : :|:  |:  | ||: ||||||| ||:|
Db          464 GVGTGTKALLETAQF-RQLQIAMHTDIQALEESISALEKSLTSLSEVVLQNRRGLDILFL 522

Qy          392 KRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWM---PW- 447
                : || |  | || |:| : : :| : : ::|:|:: | :   : : |     |    ||
Db          523 QEGGLCAALKEECCFYADHTGLVRDNMAKLRERLKQRQQLFDSQQGW--FEGWFNKSPWF 580

Qy          448 ---VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
                   :   :|||  |:|:|| |||| | ||:||  ||  |:  ::  |
Db          581 TTLISSIMGPLLILLLILLLGPCILNRLVQFVKDRISVVQALILTQ 626

RESULT 30
VCMVHL
env polyprotein precursor - HoMuLV murine leukemia virus
N;Contains: knob protein gp76; spike protein p15E
C;Species: HoMuLV murine leukemia virus
A;Note: host Mus hortulanus (European mouse)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 09-Jul-2004
C;Accession: B32594
R;Voytek, P.; Kozak, C.A.
Virology 173, 58-67, 1989
A;Title: Nucleotide sequence and mode of transmission of the wild mouse ecotropic virus,
A;Reference number: A32594; MUID:90051094; PMID:2554579
A;Accession: B32594
A;Molecule type: DNA
A;Residues: 1-666 <VOY>
A;Cross-references: UNIPROT:P21436
A;Note: the authors translated the codon CTT for residue 451 as Pro
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-466/Product: knob protein gp76 #status predicted <KPG>
F;467-666/Product: spike protein p15E #status predicted <SPP>
F;473-489/Domain: transmembrane #status predicted <TM1>
F;607-623/Domain: transmembrane #status predicted <TM2>
F;42,197,290,324,356,363,431,599/Binding site: carbohydrate (Asn) (covalent) #status prec

  Query Match             14.4%;  Score 380.5;  DB 1;  Length 666;
  Best Local Similarity   24.3%;  Pred. No. 3.5e-21;
  Matches  146;  Conservative   83;  Mismatches  206;  Indels  165;  Gaps   27;

Qy           15 PPFALTAPPPCCCTTSSSPYQEFLXRTRLPG---NIDAP-SYRSLSKGNST--------F 62
                ||::    || |||| ::|          ||   : : | :|  |:   ||
Db           92 PPYSKPPGPP-CCTTDNNP----------PGCSRDCNGPLTY--LTPRCSTAWNRLKLVL 138

Qy           63 TAH-----------THMP---RNC------YNSATLCMHANTHYWTGK------MINPSC 96
                | |            | |   |||      | :   |      ||          :: :
Db          139 TTHHLNQGFYVCPGPHRPRHARNCGGPDDFYCAHWGCETTGQAYWKPSSSWDYIRVSNNA 198

Qy           97 PGGLGATVC---------------------WT------------------YFTHTSMSD 116
                      | |                      ||                  :     :||
Db          199 SSSDATTACKNNNWCSPLAISFTDPGKRATSWTSGFTWGLRLYISGHPGLIFGVRLKISD 258

Qy          117 GGG---------IQGQAREKQVKEA----ISQLTRGHSTPS-----PYKGLVLSKLHETL 158
                 |          :  |    | : |     | ||:|  |||     | :|
Db          259 LGPRVPIGPNPVLSEQRPPSQPEPARLPPSSNLTQG-GTPSAPTGPPQEG---------- 307

Qy          159 RTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSV 218
                 |  ||: |       |:  |      ||:||      |  : |   ::| :|
Db          308 -TGDRLLDLVQGAYQALNATSPDKTQECWLCLVSSPPYYEGVAVVGPYSNHTT------- 359

Qy          219 LVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS------------ 266
                   |   : :  |   |: |     :    ||      ||  :
Db          360 --APANCSADSQHKLTLSEVT-GKPLPRKGSQ-----DPPGPVQYHSGARQKYSLSGGSR 411

Qy          267 GIFFVCGTSAYHCLNGS----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPI-- 320
                |  : | |    ||: :    :   | |  | | :| ::   :|  |  :   :| |:
Db          412 GTMWACNTGLTPCLSTAVLNLTTDYCVLVELWPRVTYHSLDFVYRQVEGRTRYQREPVSL 471

Qy          321 -LPFVIRAGVLGRLGTGIGSITTS---TQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLA 376
                 |  ::    :| :  |:|: |::   || : :|   |  |:::|  |:  |:  | ||:
Db          472 TLALLLGGLTMGGIAAGVGTGTSALVKTQQFEQLHAAIQADLKEVESSITNLEKSLTSLS 531

Qy          377 AVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTE 436
                 ||||||| ||||  ::|| |  | || |:| : : :| : : ::|:|:  | :  :  :
Db          532 EVVLQNRRGLDLLFLEKGGLCAALKEECCFYADHTGLVRDSMAKLRERLNQRQKLFEAGQ 591

Qy          437 RW--GLLSQWMPW----VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
                 |  || ::  ||    :   :|||  |:|:|:||||| | ||:||  ||  |:  :: |
Db          592 GWFEGLFNR-SPWLTTLISTIMGPLIILLLILMFGPCILNRLVQFVKDRISVVQALVLTQ 650

Search completed: January 22, 2005, 22:13:31
Job time : 42 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         January 22, 2005, 21:08:36 ; Search time 137 Seconds
                                             (without alignments)
                                             2070.509 Million cell updates/sec

Title:          US-09-319-156B-10
Perfect score:  2634
Sequence:       1 MALPYHTFLFTVLLPPFALT..........KFVSSRIEAVKLQMVLQMEP 493

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters:      1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       UniProt_02:*
                 1:  uniprot_sprot:*
                 2:  uniprot_trembl:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                   %
Result           Query
  No.   Score    Match Length DB ID                       Description
-----------------------------------------------------------------------------
    1    2613    99.2    542  2  Q991W9                   Q991w9 multiple sc
    2    2335    88.6    538  1  ENW1_GORGO               P61561 gorilla gor
    3    2323    88.2    538  1  ENW1_PANTR               P61563 pan troglod
    4    2314    87.9    538  1  ENW1_HUMAN               Q9uqf0 h herv-w_7q
    5    2313    87.8    538  1  ENW1_HYLPI               P61562 hylobates p
    6    2310    87.7    538  1  ENW1_PONPY               P61564 pongo pygma
    7    1528    58.0    319  2  Q96L59                   Q96159 homo sapien
    8    1522    57.8    319  2  Q96L60                   Q96160 homo sapien
    9    1515    57.5    319  2  Q96L61                   Q96161 homo sapien
   10     932    35.4    180  2  Q96TB1                   Q96tb1 homo sapien
   11     870    33.0    180  2  Q96TB4                   Q96tb4 homo sapien
   12     851    32.3    180  2  Q96TB5                   Q96tb5 homo sapien
   13   842.5    32.0    179  2  Q96TB3                   Q96tb3 homo sapien
   14     830    31.5    180  2  Q96TB2                   Q96tb2 homo sapien
   15     818    31.1    180  2  Q96TB6                   Q96tb6 homo sapien
   16     812    30.8    180  2  Q96TB7                   Q96tb7 homo sapien
   17   797.5    30.3    179  2  Q96TB8                   Q96tb8 homo sapien
   18     588    22.3    586  2  Q9IGU2                   Q9igu2 fowlpox vir
   19     580    22.0    563  1  ENV_BAEVM                P10269 baboon endo
   20   567.5    21.5    580  2  Q9WPZ9                   Q9wpz9 simian retr
   21   562.5    21.4    574  1  ENV_SRV2                 P51515 simian retr
   22   559.5    21.2    564  2  Q98654                   Q98654 rd114 retro
   23   559.5    21.2    567  1  ENV_AVISN                P31796 avian splee
   24     556    21.1    587  1  ENV_SRV1                 P04027 simian retr
   25   554.5    21.1    580  2  Q9WQ03                   Q9wq03 simian retr
   26   553.5    21.0    574  1  ENV_SRV2R                P51520 simian retr
   27     546    20.7    586  1  ENV_MPMV                 P07575 simian maso
   28     546    20.7    586  2  AAC82575                 Aac82575 simian ma
   29   534.5    20.3    618  2  Q8BI41                   Q8bi41 mus musculu
   30   524.5    19.9    620  2  Q9GLF7                   Q9glf7 trichosurus
   31   514.5    19.5    584  1  ENH1_HUMAN               Q9n2k0 h herv-h_2q
   32   481.5    18.3    538  1  EFR1_CALJA               P61553 callithrix
   33   477.5    18.1    575  1  ENV_SMRVH                P21412 squirrel mo
   34     477    18.1    537  1  EFR1_MACFA               P61556 macaca fasc
   35   463.5    17.6    582  1  ENV_AVIRE                P03399 avian retic
   36     463    17.6    538  1  EFR1_HYLML               P61555 hylobates m
   37     460    17.5    538  1  EFR1_HUMAN               P60508 homo sapien
   38     460    17.5    538  2  AAH68585                 Aah68585 homo sapi
   39   459.5    17.4    538  1  EFR1_PONPY               P61558 pongo pygma
   40     459    17.4    538  1  EFR1_PANTR               P61557 pan troglod
   41     454    17.2    538  1  EFR1_GORGO               P61554 gorilla gor
   42   445.5    16.9    563  1  ENH2_HUMAN               Q9n2j9 homo sapien
   43     434    16.5    555  1  ENH3_HUMAN               Q9n2j8 homo sapien
   44   428.5    16.3    584  1  EFC1_HUMAN               P60507 homo sapien
   45   421.5    16.0    584  1  EFC1_PANTR               Q8mib6 pan troglod

                                  ALIGNMENTS

RESULT 1
Q991W9
ID   Q991W9      PRELIMINARY;      PRT;   542 AA.
AC   Q991W9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Recombinant envelope protein.
GN   Name=env;
OS   Multiple sclerosis associated retrovirus element.
OC   Viruses; Retroid viruses; Retroviridae.
OX   NCBI_TaxID=89382;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21425392; PubMed=11531410;
RA   Perron H., Jouvin-Marche E., Michel M., Ounanian-Paraz A., Camelo S.,
RA   Dumon A., Jolivet-Reynaud C., Marcel F., Souillet Y., Borel E.,
RA   Gebuhrer L., Santoro L., Marcel S., Seigneurin J.M., Marche P.N.,
RA   Lafon M.;
RT   "Multiple sclerosis retrovirus particles and recombinant envelope
RT   trigger an abnormal immune response in vitro, by inducing polyclonal
RT   Vbeta16 T-lymphocyte activation.";
RL   Virology 287:321-332(2001).
DR   EMBL; AF331500; AAK18189.1; -.
DR   HSSP; P03385; 1MOF.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
SQ   SEQUENCE   542 AA;  60624 MW;  AAF3A8322E03C9EE CRC64;

  Query Match             99.2%;  Score 2613;  DB 2;  Length 542;
  Best Local Similarity   98.8%;  Pred. No. 6.7e-206;
  Matches  487;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLWRTRLPGNIDAPSYRSLSKGNS 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ||||||||||||||||||||||||||| |||||||||:| |||||||||| |||||||||
Db        421 IRDRIQCRAEELQNTERWGLLSQWMPWTLPFLGPLAAIIFLLLFGPCIFNFLVKFVSSRI 480

Qy        481 EAVKLQMVLQMEP 493
              ||||||:||||||
Db        481 EAVKLQIVLQMEP 493


RESULT 2
ENW1_GORGO
ID   ENW1_GORGO     STANDARD;      PRT;   538 AA.
AC   P61561;
DT   05-JUL-2004 (Rel. 44, Created)
DT   05-JUL-2004 (Rel. 44, Last sequence update)
DT   01-OCT-2004 (Rel. 45, Last annotation update)
DE   ERV-W1 provirus ancestral Env polyprotein precursor (Envelope
DE   polyprotein) (ERVWE1 envelope protein) (Syncytin) [Contains: Surface
DE   protein (SU); Transmembrane protein (TM)].
GN   Name=ERVWE1;
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX   NCBI_TaxID=9595;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14757826; DOI=10.1073/pnas.0305763101;
RA   Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B.,
RA   Lucotte G., Duret L., Mandrand B.;
RT   "The endogenous retroviral locus ERVWE1 is a bona fide gene involved
RT   in hominoid placental physiology.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:1731-1736(2004).
CC   -!- FUNCTION: Retroviral envelope proteins mediate receptor
CC       recognition and membrane fusion during early infection. Endogenous
CC       envelope proteins may have kept, lost or modified their original
CC       function during evolution. This endogenous envelope protein has
CC       retained its original fusogenic properties. In vitro it fuses
CC       cells expressing the type D mammalian retrovirus receptors SLC1A4
CC       and SLC1A5.
CC   -!- FUNCTION: SU mediates receptor recognition (By similarity).
CC   -!- FUNCTION: TM anchors the envelope heterodimer to the viral
CC       membrane through one transmembrane domain. The other transmembrane
CC       domain, called fusion peptide, mediates fusion of the viral
CC       membrane with the target cell membrane (By similarity).
CC   -!- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a
CC       heterodimer. SU and TM are attached by noncovalent interactions or
CC       by a labile interchain disulfide bond (By similarity).
CC   -!- DOMAIN: The CKS-17 immunosuppressive domain is present in many
CC       retroviral envelope proteins. As a synthetic peptide, it inhibits
CC       immune function in vitro and in vivo (By similarity).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield the mature SU and
CC       TM proteins (By similarity).
CC   -!- PTM: The CXXC motif is highly conserved across a broad range of
CC       retroviral envelope proteins. It is thought to participate to the
CC       formation of a labile disulfide bond possibly with the CX6CC motif
CC       present in the transmembrane protein (By similarity).
CC   -!- MISCELLANEOUS: Ortholog of the human HERV-W_7q21.1 envelope
CC       protein.
CC   -!- SIMILARITY: Belongs to the gamma type-C retroviral envelope
CC       protein family. ERV class-I W subfamily.
CC   -!- CAUTION: CKS-17 sequence does not match the minimal active
CC       concensus.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AY101588; AAM68167.1; -.
DR   EMBL; AY101589; AAM68168.1; -.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Cleavage on pair of basic residues; Envelope protein; ERV;
KW   Glycoprotein; Polyprotein; Signal; Transmembrane;
KW   Transposable element.
FT   SIGNAL        1     20       Potential.
FT   CHAIN        21    538       HERV-W1 provirus ancestral Env
FT                                polyprotein.
FT   CHAIN        21    317       Surface protein (By similarity).
FT   CHAIN       318    538       Transmembrane protein (By similarity).
FT   TRANSMEM    320    340       Fusion peptide (Potential).
FT   TRANSMEM    444    464       Potential.
FT   SITE        186    189       CXXC (By similarity).
FT   SITE        317    318       Cleavage (By similarity).
FT   SITE        380    396       CKS-17 (By similarity).
FT   SITE        397    405       CX6CC (By similarity).
FT   CARBOHYD    169    169       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    208    208       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    214    214       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    234    234       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    281    281       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    409    409       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   538 AA;  59939 MW;  95D4393FEDE196F3 CRC64;

  Query Match             88.6%;  Score 2335;  DB 1;  Length 538;
  Best Local Similarity   89.2%;  Pred. No. 4.6e-183;
  Matches  439;  Conservative   16;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              ||||||  ||||||| | ||||||| | |||||||||| | | ||||||||:|| |||
Db          1 MALPYHILLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMRRPGNIDAPSHRSFSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||||:
Db         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTHTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| ||| |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHSTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:|||| ||||:||||||:|||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLDFRPYVSIPVPEEWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||||||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTIDTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|| || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVVSKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:|||||||||||||||| |||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
```

```
              |||||||  :|:
Db       481 EAVKLQMEPKMQ 492

RESULT 3
ENW1_PANTR
ID   ENW1_PANTR     STANDARD;      PRT;   538 AA.
AC   P61563;
DT   05-JUL-2004 (Rel. 44, Created)
DT   05-JUL-2004 (Rel. 44, Last sequence update)
DT   01-OCT-2004 (Rel. 45, Last annotation update)
DE   ERV-W1 provirus ancestral Env polyprotein precursor (Envelope
DE   polyprotein) (ERVWE1 envelope protein) (Syncytin) [Contains: Surface
DE   protein (SU); Transmembrane protein (TM)].
GN   Name=ERVWE1;
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14757826; DOI=10.1073/pnas.0305763101;
RA   Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B.,
RA   Lucotte G., Duret L., Mandrand B.;
RT   "The endogenous retroviral locus ERVWE1 is a bona fide gene involved
RT   in hominoid placental physiology.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:1731-1736(2004).
CC   -!- FUNCTION: Retroviral envelope proteins mediate receptor
CC       recognition and membrane fusion during early infection. Endogenous
CC       envelope proteins may have kept, lost or modified their original
CC       function during evolution. This endogenous envelope protein has
CC       retained its original fusogenic properties. In vitro it fuses
CC       cells expressing the type D mammalian retrovirus receptors SLC1A4
CC       and SLC1A5.
CC   -!- FUNCTION: SU mediates receptor recognition (By similarity).
CC   -!- FUNCTION: TM anchors the envelope heterodimer to the viral
CC       membrane through one transmembrane domain. The other transmembrane
CC       domain, called fusion peptide, mediates fusion of the viral
CC       membrane with the target cell membrane (By similarity).
CC   -!- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a
CC       heterodimer. SU and TM are attached by noncovalent interactions or
CC       by a labile interchain disulfide bond (By similarity).
CC   -!- DOMAIN: The CKS-17 immunosuppressive domain is present in many
CC       retroviral envelope proteins. As a synthetic peptide, it inhibits
CC       immune function in vitro and in vivo (By similarity).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield the mature SU and
CC       TM proteins (By similarity).
CC   -!- PTM: The CXXC motif is highly conserved across a broad range of
CC       retroviral envelope proteins. It is thought to participate to the
CC       formation of a labile disulfide bond possibly with the CX6CC motif
CC       present in the transmembrane protein (By similarity).
CC   -!- MISCELLANEOUS: Ortholog of the human HERV-W_7q21.1 envelope
CC       protein.
CC   -!- SIMILARITY: Belongs to the gamma type-C retroviral envelope
CC       protein family. ERV class-I W subfamily.
CC   -!- CAUTION: CKS-17 sequence does not match the minimal active
CC       concensus.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AY101586; AAM68165.1; -.
DR   EMBL; AY101587; AAM68166.1; -.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Cleavage on pair of basic residues; Envelope protein; ERV;
KW   Glycoprotein; Polyprotein; Signal; Transmembrane;
```

```
KW   Transposable element.
FT   SIGNAL        1     20       Potential.
FT   CHAIN        21    538       HERV-W1 provirus ancestral Env
FT                                polyprotein.
FT   CHAIN        21    317       Surface protein (By similarity).
FT   CHAIN       318    538       Transmembrane protein (By similarity).
FT   TRANSMEM    320    340       Fusion peptide (Potential).
FT   TRANSMEM    444    464       Potential.
FT   SITE        186    189       CXXC (By similarity).
FT   SITE        317    318       Cleavage (By similarity).
FT   SITE        380    396       CKS-17 (By similarity).
FT   SITE        397    405       CX6CC (By similarity).
FT   CARBOHYD    169    169       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    208    208       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    214    214       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    234    234       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    242    242       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    281    281       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    409    409       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   538 AA;  59981 MW;  A5DB5C51EF33AB0C CRC64;

  Query Match             88.2%;  Score 2323;  DB 1;  Length 538;
  Best Local Similarity   88.6%;  Pred. No. 4.5e-182;
  Matches  436;  Conservative   18;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
             |||||| |||||||| | ||||||| | |||||||||| | : ||||||||||| |||  
Db         1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSFSKGTP 60

Qy        61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
             |||||||||||||:||||||||||||||||:|||||||||| |||||||||| ||||||:
Db        61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKIINPSCPGGLGVTVCWTYFTHTGMSDGGGV 120

Qy       121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
             | ||||| ||| |||||| ||   ||||| ||||||||||||||||||||||| ||||||
Db       121 QDQAREKHVKEVISQLTRVHSASRPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy       181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
             |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db       181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy       241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
             ||| |||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db       241 SNTTDTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy       301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
             |||||||:|| || ||||||||||| ||||| |||||| |||||||||||||||:||||:
Db       301 TEQDLYNYVVSKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy       361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
             | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db       361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy       421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
             |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db       421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy       481 EAVKLQMVLQME 492
             |||||||  :|:
Db       481 EAVKLQMEPKMQ 492


RESULT 4
ENW1_HUMAN
ID   ENW1_HUMAN     STANDARD;      PRT;   538 AA.
AC   Q9UQF0; O95244; O95245; Q8NHY7; Q9NRZ2; Q9NZG3;
DT   05-JUL-2004 (Rel. 44, Created)
DT   05-JUL-2004 (Rel. 44, Last sequence update)
DT   01-OCT-2004 (Rel. 45, Last annotation update)
DE   HERV-W_7q21.1 provirus ancestral Env polyprotein precursor (Envelope
DE   polyprotein) (HERV-7q Envelope protein) (HERV-W envelope protein)
DE   (Syncytin) (Syncytin 1) (Enverin) (Env-W) [Contains: Surface protein
DE   (SU); Transmembrane protein (TM)].
```

```
GN   Name=ERVWE1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT ASN-307.
RC   TISSUE=Placenta;
RX   MEDLINE=21985840; PubMed=11990458;
RA   Alliel P.M., Perin J.-P., Goudou D., Bitoun M., Robert B., Rieger F.;
RT   "The HERV-W/7q family in the human genome. Potential for protein
RT   expression and gene regulation.";
RL   Cell. Mol. Biol. 48:213-217(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=99099005; PubMed=9882319;
RA   Blond J.-L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
RA   Mandrand B., Mallet F.;
RT   "Molecular characterization and placental expression of HERV-W, a new
RT   human endogenous retrovirus family.";
RL   J. Virol. 73:1175-1185(1999).
RN   [3]
RP   SEQUENCE FROM N.A., TISSUE SPECIFICITY, FUNCTION, AND VARIANT ASN-307.
RC   TISSUE=Testis;
RX   MEDLINE=20155476; PubMed=10693809; DOI=10.1038/35001608;
RA   Sha M., Lee X., Li X.-P., Veldman G.M., Finnerty H., Racie L.,
RA   LaVallie E., Tang X.-Y., Edouard P., Howes S., Keith J.C. Jr.,
RA   McCoy J.M.;
RT   "Syncytin is captive retroviral envelope protein involved in human
RT   placental morphogenesis.";
RL   Nature 403:785-789(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20284713; PubMed=10826480;
RA   Voisset C., Bouton O., Bedin F., Duret L., Mandrand B., Mallet F.,
RA   Paranhos-Baccala G.;
RT   "Chromosomal distribution and coding capacity of the human endogenous
RT   retrovirus HERV-W family.";
RL   AIDS Res. Hum. Retroviruses 16:731-740(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;
RA   Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,
RA   Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,
RA   Wylie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,
RA   Fewell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,
RA   Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,
RA   Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,
RA   Ozersky P., Bielicki L., Scott K., Holmes A., Harkins R., Harris A.,
RA   Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,
RA   Kozlowicz-Reilly A., Leonard S., Rohlfing T., Rock S.M.,
RA   Tin-Wollam A.-M., Abbott A., Minx P., Maupin R., Strowmatt C.,
RA   Latreille P., Miller N., Johnson D., Murray J., Woessner J.P.,
RA   Wendl M.C., Yang S.-P., Schultz B.R., Wallis J.W., Spieth J.,
RA   Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.E., Cook L.L.,
RA   Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,
RA   Clifton S.W., Chissoe S.L., Marra M.A., Raymond C., Haugen E.,
RA   Gillett W., Zhou Y., James R., Phelps K., Iadanoto S., Bubb K.,
RA   Simms E., Levy R., Clendenning J., Kaul R., Kent W.J., Furey T.S.,
RA   Baertsch R.A., Brent M.R., Keibler E., Flicek P., Bork P., Suyama M.,
RA   Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,
RA   Eddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,
RA   Waterston R.H., Wilson R.K.;
RT   "The DNA sequence of human chromosome 7.";
RL   Nature 424:157-164(2003).
RN   [6]
RP   FUNCTION, AND TISSUE SPECIFICITY.
RX   PubMed=10708449;
RA   Blond J.-L., Lavillette D., Cheynet V., Bouton O., Oriol G.,
RA   Chapel-Fernandes S., Mandrand B., Mallet F., Cosset F.L.;
RT   "An envelope glycoprotein of the human endogenous retrovirus HERV-W is
RT   expressed in the human placenta and fuses cells expressing the type D
RT   mammalian retrovirus receptor.";
RL   J. Virol. 74:3321-3329(2000).
RN   [7]
RP   FUNCTION.
RX   PubMed=11238877; DOI=10.1128/JVI.75.7.3488-3489.2001;
RA   An D.S., Xie Y.-M., Chen I.S.Y.;
RT   "Envelope gene of the human endogenous retrovirus HERV-W encodes a
RT   functional retrovirus envelope.";
RL   J. Virol. 75:3488-3489(2001).
RN   [8]
RP   FUNCTION.
RX   PubMed=11531410; DOI=10.1006/viro.2001.1045;
RA   Perron H., Jouvin-Marche E., Michel M., Ounanian-Paraz A., Camelo S.,
RA   Dumon A., Jolivet-Reynaud C., Marcel F., Souillet Y., Borel E.,
RA   Gebuhrer L., Santoro L., Marcel S., Seigneurin J.M., Marche P.N.,
RA   Lafon M.;
RT   "Multiple sclerosis retrovirus particles and recombinant envelope
RT   trigger an abnormal immune response in vitro, by inducing polyclonal
RT   Vbeta16 T-lymphocyte activation.";
RL   Virology 287:321-332(2001).
RN   [9]
RP   FUNCTION.
RX   PubMed=12050356;
RA   Lavillette D., Marin M., Ruggieri A., Mallet F., Cosset F.-L.,
RA   Kabat D.;
RT   "The envelope glycoprotein of human endogenous retrovirus type W uses
RT   a divergent family of amino acid transporters/cell surface
RT   receptors.";
RL   J. Virol. 76:6442-6452(2002).
RN   [10]
RP   FUNCTION.
RX   MEDLINE=22946171; PubMed=14557543; DOI=10.1073/pnas.2132646100;
RA   Blaise S., de Parseval N., Benit L., Heidmann T.;
RT   "Genomewide screening for fusogenic human endogenous retrovirus
RT   envelopes identifies syncytin 2, a gene conserved on primate
RT   evolution.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:13013-13018(2003).
RN   [11]
RP   FUNCTION.
RX   PubMed=12664292; DOI=10.1007/s00705-002-0960-x;
RA   Ponferrada V.G., Mauck B.S., Wooley D.P.;
RT   "The envelope glycoprotein of human endogenous retrovirus HERV-W
RT   induces cellular resistance to spleen necrosis virus.";
RL   Arch. Virol. 148:659-675(2003).
RN   [12]
RP   TISSUE SPECIFICITY.
RX   PubMed=12970426;
RA   de Parseval N., Lazar V., Casella J.-F., Benit L., Heidmann T.;
RT   "Survey of human genes of retroviral origin: identification and
RT   transcriptome of the genes with coding capacity for complete envelope
RT   proteins.";
RL   J. Virol. 77:10414-10422(2003).
RN   [13]
RP   DEVELOPMENTAL STAGE.
RX   PubMed=12620933; DOI=10.1095/biolreprod.102.013078;
RA   Smallwood A., Papageorghiou A., Nicolaides K., Alley M.K.R., Jim A.,
RA   Nargund G., Ojha K., Campbell S., Banerjee S.;
RT   "Temporal regulation of the expression of syncytin (HERV-W),
RT   maternally imprinted PEG10, and SGCE in human placenta.";
RL   Biol. Reprod. 69:286-293(2003).
RN   [14]
RP   VARIANTS ALA-129; GLN-138; ASN-307 AND PHE-477.
RX   PubMed=14757826; DOI=10.1073/pnas.0305763101;
RA   Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B.,
RA   Lucotte G., Duret L., Mandrand B.;
RT   "The endogenous retroviral locus ERVWE1 is a bona fide gene involved
RT   in hominoid placental physiology.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:1731-1736(2004).
CC   -!- FUNCTION: Retroviral envelope proteins mediate receptor
CC       recognition and membrane fusion during early infection. Endogenous
CC       envelope proteins may have kept, lost or modified their original
CC       function during evolution. This endogenous envelope protein has
CC       retained its original fusogenic properties. It can make
CC       pseudotypes with HIV-1 virions and confer infectivity. It
CC       participates to trophoblast fusion during placenta morphogenesis.
CC       In vitro it fuses cells expressing the type D mammalian retrovirus
CC       receptors SLC1A4 and SLC1A5. It can induce cellular resistance to
CC       spleen necrosis virus in vitro.
CC   -!- FUNCTION: SU mediates receptor recognition (By similarity).
CC   -!- FUNCTION: TM anchors the envelope heterodimer to the viral
CC       membrane through one transmembrane domain. The other transmembrane
CC       domain, called fusion peptide, mediates fusion of the viral
CC       membrane with the target cell membrane (By similarity).
CC   -!- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a
CC       heterodimer. SU and TM are attached by noncovalent interactions or
CC       by a labile interchain disulfide bond (By similarity).
CC   -!- TISSUE SPECIFICITY: Expressed at higher level in placenta.
CC       Expressed at lower level in adrenal, bone marrow, brain, breast,
CC       colon, kidney, ovary, prostate, skin, spleen, testis, thymus,
CC       thyroid and trachea.
CC   -!- DEVELOPMENTAL STAGE: In placenta, detected at higher level during
CC       early pregnancy and at lower level during late pregnancy.
CC   -!- DOMAIN: The CKS-17 immunosuppressive domain is present in many
CC       retroviral envelope proteins. As a synthetic peptide, it inhibits
CC       immune function in vitro and in vivo (By similarity).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield the mature SU and
CC       TM proteins (By similarity).
CC   -!- PTM: The CXXC motif is highly conserved across a broad range of
CC       retroviral envelope proteins. It is thought to participate to the
CC       formation of a labile disulfide bond possibly with the CX6CC motif
CC       present in the transmembrane protein. Isomerization of the
CC       intersubunit disulfide bond to an SU intrachain disulfide bond is
CC       thought to occur upon receptor recognition in order to allow
CC       membrane fusion (By similarity).
CC   -!- POLYMORPHISM: All variants have fusogenic properties.
CC   -!- DISEASE: Viral particles or intracellular RNA of HERV-W family
CC       members have been detected in tissue from patients with multiple
CC       sclerosis or schizophrenia.
CC   -!- MISCELLANEOUS: Orthologs in Pan troglodytes, Gorilla gorilla,
CC       Pongo pygmaeus and Hylobates moloch.
CC   -!- MISCELLANEOUS: HERV-W family subgenomic RNAs have been observed.
CC   -!- MISCELLANEOUS: This provirus is intergenic, the closest flanking
CC       genes being ODAG and PEX1.
CC   -!- SIMILARITY: Belongs to the gamma type-C retroviral envelope
CC       protein family. HERV class-I W subfamily.
CC   -!- CAUTION: CKS-17 sequence does not match the minimal active
CC       concensus.
CC   -!- DATABASE: NAME=Human Endogenous Retrovirus Database (HERVd);

  Query Match             87.9%;  Score 2314;  DB 1;  Length 538;
  Best Local Similarity   88.4%;  Pred. No. 2.5e-181;
  Matches  435;  Conservative   19;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : ||||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||:||||:||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||  ||:|||||||||||:||||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              ||||||::|: || ||||||||||| ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360
```

```
Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:||  ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492


RESULT 5
ENW1_HYLPI
ID   ENW1_HYLPI     STANDARD;      PRT;   538 AA.
AC   P61562;
DT   05-JUL-2004 (Rel. 44, Created)
DT   05-JUL-2004 (Rel. 44, Last sequence update)
DT   01-OCT-2004 (Rel. 45, Last annotation update)
DE   ERV-W1 provirus ancestral Env polyprotein precursor (Envelope
DE   polyprotein) (ERVWE1 envelope protein) (Syncytin) [Contains: Surface
DE   protein (SU); Transmembrane protein (TM)].
GN   Name=ERVWE1;
OS   Hylobates pileatus (Pileated gibbon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
OX   NCBI_TaxID=9589;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14757826; DOI=10.1073/pnas.0305763101;
RA   Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B.,
RA   Lucotte G., Duret L., Mandrand B.;
RT   "The endogenous retroviral locus ERVWE1 is a bona fide gene involved
RT   in hominoid placental physiology.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:1731-1736(2004).
CC   -!- FUNCTION: Retroviral envelope proteins mediate receptor
CC       recognition and membrane fusion during early infection. Endogenous
CC       envelope proteins may have kept, lost or modified their original
CC       function during evolution. This endogenous envelope protein has
CC       retained its original fusogenic properties. In vitro it fuses
CC       cells expressing the type D mammalian retrovirus receptors SLC1A4
CC       and SLC1A5.
CC   -!- FUNCTION: SU mediates receptor recognition (By similarity).
CC   -!- FUNCTION: TM anchors the envelope heterodimer to the viral
CC       membrane through one transmembrane domain. The other transmembrane
CC       domain, called fusion peptide, mediates fusion of the viral
CC       membrane with the target cell membrane (By similarity).
CC   -!- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a
CC       heterodimer. SU and TM are attached by noncovalent interactions or
CC       by a labile interchain disulfide bond (By similarity).
CC   -!- DOMAIN: The CKS-17 immunosuppressive domain is present in many
CC       retroviral envelope proteins. As a synthetic peptide, it inhibits
CC       immune function in vitro and in vivo (By similarity).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield the mature SU and
CC       TM proteins (By similarity).
CC   -!- PTM: The CXXC motif is highly conserved across a broad range of
CC       retroviral envelope proteins. It is thought to participate to the
CC       formation of a labile disulfide bond possibly with the CX6CC motif
CC       present in the transmembrane protein (By similarity).
CC   -!- MISCELLANEOUS: Ortholog of the human HERV-W_7q21.1 envelope
CC       protein.
CC   -!- SIMILARITY: Belongs to the gamma type-C retroviral envelope
CC       protein family. ERV class-I W subfamily.
CC   -!- CAUTION: CKS-17 sequence does not match the minimal active
CC       concensus.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AY101592; AAM68171.1; -.
DR   EMBL; AY101593; AAM68172.1; -.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Cleavage on pair of basic residues; Envelope protein; ERV;
KW   Glycoprotein; Polyprotein; Signal; Transmembrane;
KW   Transposable element.
FT   SIGNAL        1     20       Potential.
FT   CHAIN        21    538       HERV-W1 provirus ancestral Env
FT                                polyprotein.
FT   CHAIN        21    317       Surface protein (By similarity).
FT   CHAIN       318    538       Transmembrane protein (By similarity).
FT   TRANSMEM    320    340       Fusion peptide (Potential).
FT   TRANSMEM    444    464       Potential.
FT   SITE        186    189       CXXC (By similarity).
FT   SITE        317    318       Cleavage (By similarity).
FT   SITE        380    396       CKS-17 (By similarity).
FT   SITE        397    405       CX6CC (By similarity).
FT   CARBOHYD    169    169       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    208    208       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    214    214       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    234    234       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    242    242       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    281    281       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    409    409       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   538 AA;  60100 MW;  E833A5AF7AF66BD7 CRC64;

  Query Match             87.8%;  Score 2313;  DB 1;  Length 538;
  Best Local Similarity   88.4%;  Pred. No. 3e-181;
  Matches  435;  Conservative   18;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| ||: ||||||| ||| |||:
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRTQRPGNIDAPLYRSFSKGSP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||:||| ||||||||||||| |||||||||||||||| |||||||||| ||||||:
Db         61 TFTAHTYMPRTCYNSATLCMHANTQYWTGKMINPSCPGGLGVTVCWTYFTHTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||: ||| |||||| ||||||||||||||||||||||| ||||||
Db        121 QDQAREKHVKEVISQLTQVHSTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||||||| ||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLDFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||| |||:||||||||||||| ||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTDTTNSQCIRWVTPPTRIFCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|| || ||||||||||: ||||| |||||||||||||||||||:|:|||||:
Db        301 TEQDLYNYVVSKPRNKRVPILPFVMGAGVLGALGTGIGSITTSTQFYYKLSRELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | |||||||||||||||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSLAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:|   ||||||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNIGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              ||:||||  :||
Db        481 EAIKLQMEPKME 492


RESULT 6
```

```
ENW1_PONPY
ID   ENW1_PONPY     STANDARD;      PRT;   538 AA.
AC   P61564;
DT   05-JUL-2004 (Rel. 44, Created)
DT   05-JUL-2004 (Rel. 44, Last sequence update)
DT   01-OCT-2004 (Rel. 45, Last annotation update)
DE   ERV-W1 provirus ancestral Env polyprotein precursor (Envelope
DE   polyprotein) (ERVWE1 envelope protein) (Syncytin) [Contains: Surface
DE   protein (SU); Transmembrane protein (TM)].
GN   Name=ERVWE1;
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14757826; DOI=10.1073/pnas.0305763101;
RA   Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B.,
RA   Lucotte G., Duret L., Mandrand B.;
RT   "The endogenous retroviral locus ERVWE1 is a bona fide gene involved
RT   in hominoid placental physiology.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:1731-1736(2004).
CC   -!- FUNCTION: Retroviral envelope proteins mediate receptor
CC       recognition and membrane fusion during early infection. Endogenous
CC       envelope proteins may have kept, lost or modified their original
CC       function during evolution. This endogenous envelope protein has
CC       retained its original fusogenic properties. In vitro it fuses
CC       cells expressing the type D mammalian retrovirus receptors SLC1A4
CC       and SLC1A5.
CC   -!- FUNCTION: SU mediates receptor recognition (By similarity).
CC   -!- FUNCTION: TM anchors the envelope heterodimer to the viral
CC       membrane through one transmembrane domain. The other transmembrane
CC       domain, called fusion peptide, mediates fusion of the viral
CC       membrane with the target cell membrane (By similarity).
CC   -!- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a
CC       heterodimer. SU and TM are attached by noncovalent interactions or
CC       by a labile interchain disulfide bond (By similarity).
CC   -!- DOMAIN: The CKS-17 immunosuppressive domain is present in many
CC       retroviral envelope proteins. As a synthetic peptide, it inhibits
CC       immune function in vitro and in vivo (By similarity).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield the mature SU and
CC       TM proteins (By similarity).
CC   -!- PTM: The CXXC motif is highly conserved across a broad range of
CC       retroviral envelope proteins. It is thought to participate to the
CC       formation of a labile disulfide bond possibly with the CX6CC motif
CC       present in the transmembrane protein (By similarity).
CC   -!- MISCELLANEOUS: Ortholog of the human HERV-W_7q21.1 envelope
CC       protein.
CC   -!- SIMILARITY: Belongs to the gamma type-C retroviral envelope
CC       protein family. ERV class-I W subfamily.
CC   -!- CAUTION: CKS-17 sequence does not match the minimal active
CC       concensus.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AY101590; AAM68169.1; -.
DR   EMBL; AY101591; AAM68170.1; -.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Cleavage on pair of basic residues; Envelope protein; ERV;
KW   Glycoprotein; Polyprotein; Signal; Transmembrane;
KW   Transposable element.
FT   SIGNAL        1     20       Potential.
FT   CHAIN        21    538       HERV-W1 provirus ancestral Env
FT                                polyprotein.
FT   CHAIN        21    317       Surface protein (By similarity).
FT   CHAIN        318    538       Transmembrane protein (By similarity).
FT   TRANSMEM     320    340       Fusion peptide (Potential).
FT   TRANSMEM     444    464       Potential.
FT   SITE         186    189       CXXC (By similarity).
FT   SITE         317    318       Cleavage (By similarity).
FT   SITE         380    396       CKS-17 (By similarity).
FT   SITE         397    405       CX6CC (By similarity).
FT   CARBOHYD     169    169       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD     208    208       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD     214    214       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD     234    234       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD     242    242       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD     245    245       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD     281    281       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD     409    409       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   538 AA;  59908 MW;  AA17EE08E6D62A99 CRC64;

  Query Match            87.7%;  Score 2310;  DB 1;  Length 538;
  Best Local Similarity   87.8%;  Pred. No. 5.3e-181;
  Matches  432;  Conservative   19;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| |    ||||||||||| |||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMHHPGNIDAPSYRSFSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              ||||||||||||||||||||| ||||||||||||||||||| |||||||||| ||||||:
Db         61 TFTAHTHMPRNCYNSATLCMHGNTHYWTGKMINPSCPGGLGVTVCWTYFTHTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | :|||| ||| |||||: ||| |||||| ||||||||||||||||||||||| ||||||
Db        121 QDEAREKHVKEVISQLTQVHSTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              |||||||||||| ||||:||||||||||||||||||||||||||||||||||||||||||
Db        181 QNPTNCWMCLPLAFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              ||| :||:|||||||||||::|||||||||||||||| ||||||||||||||||||||||
Db        241 SNTTNTTNSQCIRWVTPPTQVVCLPSGIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQ 360
              |||||||:|| || ||||||||||: ||||| |||||| ||||||||||||||:|||||:
Db        301 TEQDLYNYVVSKPRNKRVPILPFVMAAGVLGALGTGIGGITTSTQFYYKLSQELNGDMER 360

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              | ||||||||||||:||||||||||||||||:||||||||||| ||||||| ||||||||
Db        361 VADSLVTLQDQLNSIAAVVLQNRRALDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKE 420

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              |||||| |||||:|   ||| ||||||:|||||||||:||||||||||||||| ||||||
Db        421 IRDRIQRRAEELRNIGPWGLFSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRI 480

Qy        481 EAVKLQMVLQME 492
              |||||||  :|:
Db        481 EAVKLQMEPKMQ 492

RESULT 7
Q96L59
ID   Q96L59      PRELIMINARY;      PRT;   319 AA.
AC   Q96L59;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Envelope protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jern P., Karlsson D., Blomberg J.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY050301; AAL11495.1; -.
DR   GO; GO:0019031; C:viral envelope; IEA.
KW   Envelope protein.
FT   NON_TER     319    319
SQ   SEQUENCE   319 AA;  35784 MW;  7D288CCCD1B68999 CRC64;

  Query Match            58.0%;  Score 1528;  DB 2;  Length 319;
  Best Local Similarity   87.8%;  Pred. No. 4.9e-117;
  Matches  280;  Conservative    8;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| |||||||||||||||||||| || ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTALTGLHEVSA 180

Qy        181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
              ||||| ||||||||||||| ||||||||||||| ||||||||| ||:|| ||||||||||
Db        181 QNPTNSWMCLPLHFRPYISTPVPEQWNNFSTEIITTSVLVGPLASNVEIIHTSNLTCVKF 240

Qy        241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
              |||||||:|||||||||||||||||||||||||||||||||  |||||||||||||||||
Db        241 SNTIDTTNSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNSFSESMCFLSFLVPPMTIY 300

Qy        301 TEQDLYNHVVPKPHNKRVP 319
              ||||||::|: || |||||
Db        301 TEQDLYSYVISKPRNKRVP 319

RESULT 8
Q96L60
ID   Q96L60      PRELIMINARY;      PRT;   319 AA.
AC   Q96L60;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Envelope protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jern P., Karlsson D., Blomberg J.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY050300; AAL11494.1; -.
DR   GO; GO:0019031; C:viral envelope; IEA.
KW   Envelope protein.
FT   NON_TER     319    319
SQ   SEQUENCE   319 AA;  35763 MW;  67822669D1B6899D CRC64;

  Query Match            57.8%;  Score 1522;  DB 2;  Length 319;
  Best Local Similarity   87.5%;  Pred. No. 1.5e-116;
  Matches  279;  Conservative    8;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
              |||||| |||||||| | ||||||| | |||||||||| | : |||||||||||||||
Db          1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy         61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
              |||||||||||||:||||| ||||||||||||||||||||| |||||||| | ||||||:
Db         61 TFTAHTHMPRNCYHSATLCXHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy        121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
              | ||||| ||| |||||| | | |||||| |||||||||||||||||||| || ||||||
Db        121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTALTGLHEVSA 180
```

```
Qy           181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
                 ||||| ||||||||||||| ||||||||||||| ||||||||| ||:|| ||||||||||
Db           181 QNPTNSWMCLPLHFRPYISTPVPEQWNNFSTEIITTSVLVGPLASNVEIIHTSNLTCVKF 240

Qy           241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
                 |||||||:|||||||||||||||||||||||||||||||||  |||||||||||||||||
Db           241 SNTIDTTNSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNSFSESMCFLSFLVPPMTIY 300

Qy           301 TEQDLYNHVVPKPHNKRVP 319
                 ||||||::|: || |||||
Db           301 TEQDLYSYVISKPRNKRVP 319


RESULT 9
Q96L61
ID   Q96L61       PRELIMINARY;      PRT;   319 AA.
AC   Q96L61;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Envelope protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jern P., Karlsson D., Blomberg J.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY050299; AAL11493.1; -.
DR   GO; GO:0019031; C:viral envelope; IEA.
KW   Envelope protein.
FT   NON_TER     319    319
SQ   SEQUENCE   319 AA;  35655 MW;  7D3397CCCAAD8999 CRC64;

  Query Match             57.5%;  Score 1515;  DB 2;  Length 319;
  Best Local Similarity   87.5%;  Pred. No. 5.8e-116;
  Matches  279;  Conservative    8;  Mismatches   32;  Indels    0;  Gaps    0;

Qy             1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNS 60
                 |||||| |||||||| | ||||||| | | |||||||||| | : |||||||||||||||
Db             1 MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRPGNIDAPSYRSLSKGTP 60

Qy            61 TFTAHTHMPRNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSDGGGI 120
                 |||||||||||||:||||||||||||||||||||||||||| |||||||| | ||||||:
Db            61 TFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFTQTGMSDGGGV 120

Qy           121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSA 180
                 | ||||| ||| |||||| | | |||||| |||||||||||||||||||| || ||||||
Db           121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTALTGLHEVSA 180

Qy           181 QNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKF 240
                 ||||| ||||||||||||| ||||||||||||| ||||||||| ||:|| ||||||||||
Db           181 QNPTNSWMCLPLHFRPYISTPVPEQWNNFSTEIITTSVLVGPLASNVEIIHTSNLTCVKF 240

Qy           241 SNTIDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
                 |||||||:||||| |||||||||||||||||||||||||||  |||||||||||||||||
Db           241 SNTIDTTNSQCIRGVTPPTRIVCLPSGIFFVCGTSAYHCLNSFSESMCFLSFLVPPMTIY 300

Qy           301 TEQDLYNHVVPKPHNKRVP 319
                 ||||||::|: || |||||
Db           301 TEQDLYSYVISKPRNKRVP 319


RESULT 10
Q96TB1
ID   Q96TB1       PRELIMINARY;      PRT;   180 AA.
AC   Q96TB1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein (Fragment).
```

```
GN   Name=env;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21268768; PubMed=11375061;
RA   Kim H.-S., Lee W.-H.;
RT   "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT   identification, and phylogeny.";
RL   AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kim H.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051010; BAB47562.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     180    180
SQ   SEQUENCE   180 AA;  19759 MW;  C4D3C9F72388A0C3 CRC64;

  Query Match             35.4%;  Score 932;  DB 2;  Length 180;
  Best Local Similarity   99.4%;  Pred. No. 2.3e-68;
  Matches  179;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy           207 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS 266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db             1 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS 60

Qy           267 GIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            61 GIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 120

Qy           327 AGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRAL 386
                 |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db           121 AGVLGRLGTGIGGITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRAL 180


RESULT 11
Q96TB4
ID   Q96TB4       PRELIMINARY;      PRT;   180 AA.
AC   Q96TB4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein (Fragment).
GN   Name=env;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21268768; PubMed=11375061;
RA   Kim H.-S., Lee W.-H.;
RT   "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT   identification, and phylogeny.";
RL   AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kim H.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051007; BAB47559.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     180    180
SQ   SEQUENCE   180 AA;  20048 MW;  F2B7B2FAF7CE40C3 CRC64;

  Query Match             33.0%;  Score 870;  DB 2;  Length 180;
  Best Local Similarity   92.8%;  Pred. No. 2.8e-63;
  Matches  167;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        207 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS 266
              |||||||||||:|||||||||||||||||| ||||||||||:|||||||||||:||||||
Db          1 NNFSTEINTTSILVGPLVSNLEITHTSNLTRVKFSNTIDTTNSQCIRWVTPPTQIVCLPS 60

Qy        267 GIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
              |||||||||||||||| ||||||||||| ||||||||||||:||||| ||||||||||||
Db         61 GIFFVCGTSAYHCLNGFSESMCFLSFLVHPMTIYTEQDLYNYVVPKPRNKRVPILPFVIR 120

Qy        327 AGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRAL 386
              ||:|||||||:||||||||||||||||:||||||:| |||||||||||||||||||||||
Db        121 AGMLGRLGTGMGSITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRAL 180


RESULT 12
Q96TB5
ID   Q96TB5      PRELIMINARY;      PRT;   180 AA.
AC   Q96TB5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein (Fragment).
GN   Name=env;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21268768; PubMed=11375061;
RA   Kim H.-S., Lee W.-H.;
RT   "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT   identification, and phylogeny.";
RL   AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kim H.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051004; BAB47558.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     180    180
SQ   SEQUENCE   180 AA;  19648 MW;  8D2F33813F16090B CRC64;

  Query Match             32.3%;  Score 851;  DB 2;  Length 180;
  Best Local Similarity   91.1%;  Pred. No. 1e-61;
  Matches  164;  Conservative    7;  Mismatches    9;  Indels    0;  Gaps    0;

Qy        207 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS 266
              |||||||||||||||||||||||||||||||||||||  ||:|||||||||||:||||||
Db          1 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPS 60

Qy        267 GIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
              ||||||||||| ||||||||||||||||||||||||||||::|: || |||||||||||
Db         61 GIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIG 120

Qy        327 AGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRAL 386
              ||||| |||||| ||||||||||||||:||||||:| |||||||||||||||||||||||
Db        121 AGVLGALGTGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRAL 180
```

```
RESULT 13
Q96TB3
ID   Q96TB3      PRELIMINARY;      PRT;   179 AA.
AC   Q96TB3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein (Fragment).
GN   Name=env;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21268768; PubMed=11375061;
RA   Kim H.-S., Lee W.-H.;
RT   "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT   identification, and phylogeny.";
RL   AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kim H.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051008; BAB47560.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     179    179
SQ   SEQUENCE   179 AA;  19471 MW;  70DF3E5658B8E667 CRC64;

  Query Match             32.0%;  Score 842.5;  DB 2;  Length 179;
  Best Local Similarity   90.6%;  Pred. No. 5.1e-61;
  Matches  163;  Conservative    6;  Mismatches   10;  Indels    1;  Gaps    1;

Qy        207 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS 266
              ||||||||||||||||| |||||||||||||||||||||| :||||||||||||||||||
Db          1 NNFSTEINTTSVLVGPL-SNLEITHTSNLTCVKFSNTIDTANSQCIRWVTPPTRIVCLPS 59

Qy        267 GIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
              ||||||||||||||||||||:||||||| || |||||||||||:||| |||||||||||
Db         60 GIFFVCGTSAYHCLNGSSESVCFLSFLVAPMPIYTEQDLYNHVIPKPRNKRVPILPFVIG 119

Qy        327 AGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRAL 386
              ||||| : |||| ||||||||||||||:|||||| | |:|||||||||||||||||||||
Db        120 AGVLGGVATGIGGITTSTQFYYKLSQELNGDMEWVADTLVTLQDQLNSLAAVVLQNRRAL 179


RESULT 14
Q96TB2
ID   Q96TB2      PRELIMINARY;      PRT;   180 AA.
AC   Q96TB2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein (Fragment).
GN   Name=env;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21268768; PubMed=11375061;
RA   Kim H.-S., Lee W.-H.;
RT   "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT   identification, and phylogeny.";
```

```
RL   AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kim H.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051009; BAB47561.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     180    180
SQ   SEQUENCE   180 AA;  19709 MW;  4C8F196C9C788497 CRC64;

  Query Match             31.5%;  Score 830;  DB 2;  Length 180;
  Best Local Similarity   88.3%;  Pred. No. 5.4e-60;
  Matches  159;  Conservative    7;  Mismatches   14;  Indels    0;  Gaps    0;

Qy        207 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS 266
              |||||||||||:|:|||||||||||||||||||||||||| :|||||||: |||||||||
Db          1 NNFSTEINTTSILIGPLVSNLEITHTSNLTCVKFSNTIDTPNSQCIRWVSHPTRIVCLPS 60

Qy        267 GIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
              ||||||||||| |  |||||:|||||||||||||||||||||||||| ||||||| ||:
Db         61 GIFFVCGTSAYRCFKGSSESICFLSFLVPPMTIYTEQDLYNHVVPKPRNKRVPILTFVVG 120

Qy        327 AGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRAL 386
              || || |||||| ||||||||||||||:||||| | ||||||||||||| ||||||||||
Db        121 AGGLGGLGTGIGGITTSTQFYYKLSQELNGDMEWVADSLVTLQDQLNSLVAVVLQNRRAL 180


RESULT 15
Q96TB6
ID   Q96TB6      PRELIMINARY;      PRT;   180 AA.
AC   Q96TB6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein (Fragment).
GN   Name=env;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21268768; PubMed=11375061;
RA   Kim H.-S., Lee W.-H.;
RT   "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT   identification, and phylogeny.";
RL   AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kim H.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051000; BAB47557.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     180    180
SQ   SEQUENCE   180 AA;  19656 MW;  EF8DFCBD4CAEAF3D CRC64;

  Query Match             31.1%;  Score 818;  DB 2;  Length 180;
  Best Local Similarity   88.3%;  Pred. No. 5.3e-59;
  Matches  159;  Conservative    7;  Mismatches   14;  Indels    0;  Gaps    0;

Qy        207 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS 266
```

```
              |||||||||||||||||||||||||||||||||||||||||:|||||||||||:||||||
Db          1 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTNSQCIRWVTPPTQIVCLPS 60

Qy        267 GIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
              |||||||||||||||||| || ||||| ||| |||||||||:|:|| |||||||| |||
Db         61 GIFFVCGTSAYHCLNGSSGSMWFLSFLAPPMIIYTEQDLYNYVIPKSHNKRVPILSFVIG 120

Qy        327 AGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRAL 386
              ||||  || ||| | ||||||||||||:||||: ||:|||||||||||||||| ||||||
Db        121 AGVLDGLGAGIGGIKTSTQFYYKLSQELNGDMDWVTNSLVTLQDQLNSLAAVVPQNRRAL 180


RESULT 16
Q96TB7
ID   Q96TB7      PRELIMINARY;      PRT;   180 AA.
AC   Q96TB7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein (Fragment).
GN   Name=env;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21268768; PubMed=11375061;
RA   Kim H.-S., Lee W.-H.;
RT   "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT   identification, and phylogeny.";
RL   AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kim H.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB050999; BAB47556.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     180    180
SQ   SEQUENCE   180 AA;  19984 MW;  3414BCDD74A060BA CRC64;

  Query Match             30.8%;  Score 812;  DB 2;  Length 180;
  Best Local Similarity   87.8%;  Pred. No. 1.6e-58;
  Matches  158;  Conservative    6;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        207 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS 266
              |||||||||||:||||||||||||||||||||||||||||| |||||||||||||||||
Db          1 NNFSTEINTTSILVGPLVSNLEITHTSNLTCVKFSNTIDTTKSQCIRWVTPPTRIVCLPL 60

Qy        267 GIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
              ||||| |: || |||| || ||||||||||||||||| |||:|:|||||| | |||||||
Db         61 GIFFVFGSLAYRCLNGFSEPMCFLSFLVPPMTIYTEQYLYNYVIPKPHNKTVLILPFVIR 120

Qy        327 AGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRAL 386
              ||||| | |||| ||||||||||||||: |||||||||||||||  ||||||||||:|||
Db        121 AGVLGGLSTGIGGITTSTQFYYKLSQELKGDMEQVTDSLVTLQDHHNSLAAVVLQNQRAL 180


RESULT 17
Q96TB8
ID   Q96TB8      PRELIMINARY;      PRT;   179 AA.
AC   Q96TB8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein (Fragment).
GN   Name=env;
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21268768; PubMed=11375061;
RA   Kim H.-S., Lee W.-H.;
RT   "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT   identification, and phylogeny.";
RL   AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kim H.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB050996; BAB47555.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
FT   NON_TER     179    179
SQ   SEQUENCE   179 AA;  19228 MW;  3A9FACBAB6376E74 CRC64;

  Query Match             30.3%;  Score 797.5;  DB 2;  Length 179;
  Best Local Similarity   87.8%;  Pred. No. 2.5e-57;
  Matches  158;  Conservative    6;  Mismatches   15;  Indels    1;  Gaps    1;

Qy        207 NNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTPPTRIVCLPS 266
              |||||||||||:||||||||||| |||||||||||||| | :||||| |||||||||| |
Db          1 NNFSTEINTTSILVGPLVSNLEIIHTSNLTCVKFSNTIATANSQCIRRVTPPTRIVCLTS 60

Qy        267 GIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
              ||||||||||| |||||||||||||||||||||||||:||||| ||| |||||||||||
Db         61 GIFFVCGTSAYRCLNGSSESMCFLSFLVPPMTIYTEQNLYNHVAPKPRNKRVPILPFVIG 120

Qy        327 AGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRAL 386
              ||||| |||| | || |||| ||||||:|||||:| ||||||||||||||||||||: ||
Db        121 AGVLGGLGTGPGGITASTQF-YKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNQTAL 179


RESULT 18
Q9IGU2
ID   Q9IGU2       PRELIMINARY;      PRT;   586 AA.
AC   Q9IGU2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope glycoprotein.
OS   Fowlpox virus (FPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Avipoxvirus.
OX   NCBI_TaxID=10261;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Singh P., Kim T.-J., Tripathy D.N.;
RT   "Re-emerging fowlpox: evaluation of isolates from vaccinated flocks.";
RL   Avian Pathol. 29:449-455(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Singh P., Schitzlein W.M., Tripathy D.N.;
RT   "Reticuloendotheliosis Virus Sequences Within the Genomes of Field
RT   Strains of Fowlpox Virus Display Variability.";
RL   J. Virol. 0:0-0(2003).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Schnitzlein W.M., Singh P., Srinivasan V., Tripathy D.N.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF246698; AAF81698.2; -.
DR   HSSP; P03385; 1MOF.
DR   GO; GO:0019028; C:viral capsid; IEA.
```

```
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR008965; Cellul_bind.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
SQ   SEQUENCE   586 AA;  63723 MW;  DBB221FA9FC562C3 CRC64;

  Query Match             22.3%;  Score 588;  DB 2;  Length 586;
  Best Local Similarity   32.6%;  Pred. No. 1.9e-39;
  Matches  161;  Conservative   74;  Mismatches  161;  Indels   98;  Gaps   20;

Qy         63 TAHTHMPRNCYNSATLCMHANTHYWT---------------GKMINPSCPGGLGATVCWT 107
              |  | |   ||  |  |      |:|                |::  || | :|  |||
Db        118 TIATQMHSTCYEKAQECTLLGKTYFTAILQKTKLGSYEDGPNKLLQASCTGTIGKPVCWD 177

Qy        108 YFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPS-PYKGLVLSKLHETLRTHTRLVS 166
                    :|||||     ||: |:| : :: | || ||  |  | | :         | |
Db        178 PVAPVYVSDGGGPTDMIREESVRERLEEIIR-HSYPSVQYHPLALPR--------PRGVD 228

Qy        167 LFNTT---LTRLHEV-SAQNP---TNCWMC--------------------------LPLH 193
              |   |   |   |:| :| ||    |||:|                          ||
Db        229 LDPQTSDILEATHQVLNATNPQLAENCWLCMTLGTPIPAAIPANGNVTLDGNCSLSLPFR 288

Qy        194 FRPYISIPVPEQWNNFSTEI-NTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCI 252
               :|  || |    | :: |  | | : :|         | :|  |       : ||   |
Db        289 VQPTGSIDV----NCYAGEADNRTGIPIG-------YVHFTN--CTSIQEVSNETSH--I 333

Qy        253 RWVTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVP 311
              | :|     :| | |  |||| : ||  |      :| |: :|| |:| : ::     :|
Db        334 RNLTR----LCPPPGHVFVCGNNMAYTALPNKWIGLCILASIVPDMSIISGEE----PIP 385

Qy        312 KP--------HNKRVPILPFVIRAGVLGRL---GTGIGSITTSTQFYYKLSQEINGDMEQ 360
               |        | : |  :| ::  |: |     |||:|    |   |:||| ::  |::
Db        386 LPSIEYTAGRHKRAVQFIPLLVGLGITGATLAGGTGLG---VSVHTYHKLSNQLIEDVQA 442

Qy        361 VTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE 420
              :: ::  ||||::||| ||||||| ||||||::|| || | |: |:| |:| || :|:::
Db        443 LSGTINDLQDQIDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIRK 502

Qy        421 IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRI 480
              ::: :  |   | :   |  |: ::|::|| ||||  ||| |  ||||   | : :  :|
Db        503 LQEDLIARKRALYDNPLWNGLNGFLPYLLPLLGPLFGLILFLTLGPCIIKTLTRIIHDKI 562

Qy        481 EAVK-LQMVLQMEP 493
              :||| | :| | :|
Db        563 QAVKILALVPQYKP 576


RESULT 19
ENV_BAEVM
ID   ENV_BAEVM      STANDARD;      PRT;   563 AA.
AC   P10269;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Env polyprotein precursor (Coat polyprotein) [Contains: Outer membrane
DE   protein GP70; Transmembrane protein p20E].
GN   Name=ENV;
OS   Baboon endogenous virus (strain M7).
OC   Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
OX   NCBI_TaxID=11764;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kato S., Matsuo K., Nishimura N., Takahashi N., Takano T.;
RT   "The entire nucleotide sequence of baboon endogenous virus DNA: a
RT   chimeric genome structure of murine type C and simian type D
RT   retroviruses.";
RL   Jpn. J. Genet. 62:127-137(1987).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D10032; BAA00924.1; -.
DR   EMBL; X05470; CAA29028.1; -.
DR   PIR; JT0262; VCMVM7.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Coat protein; Glycoprotein; Polyprotein; Signal; Transmembrane.
FT   SIGNAL        1     20       Potential.
FT   CHAIN        21    346       Outer membrane protein GP70 (Potential).
FT   CHAIN       347    563       Transmembrane protein p20E (Potential).
FT   CARBOHYD    113    113       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    219    219       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    229    229       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    264    264       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    282    282       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    292    292       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    306    306       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    312    312       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    321    321       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    339    339       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    469    469       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   563 AA;  61879 MW;  9573137DC4620BB7 CRC64;

  Query Match             22.0%;  Score 580;  DB 1;  Length 563;
  Best Local Similarity   29.2%;  Pred. No. 8.2e-39;
  Matches  146;  Conservative   98;  Mismatches  190;  Indels   66;  Gaps   14;

Qy         28 TTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNSTFTAHTHMPRNCYNSATLCMHANTHYW 87
              |: | | ||       | |    ||:|        :| :    :|| |   |   |  |:
Db         80 TSPSGPLQE------CPCN----SYQS--------SVHS----SCYTSYQQCRSGNKTYY 117

Qy         88 --------------------TGKMINPSCPGGLGATVCWTYFTHTSMSDGGGIQGQAREK 127
                                  | |:|   | |  | ::||:      :|||||     | |
Db        118 TATLLKTQTGGTSDVQVLGSTNKLIQSPCNGIKGQSICWSTTAPIHVSDGGGPLDTTRIK 177

Qy        128 QVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCW 187
               |:  : :: :       |  | : |: : |    : ::: | |   |   :     :||
Db        178 SVQRKLEEIHKALYPELQYHPLAIPKVRDNLMVDAQTLNILNATYNLLLMSNTSLVDDCW 237

Qy        188 MCLPLHFRPYISIPVPEQWNNFSTEI--NTTSVLVGPLVSNLEITHTSNLTCVKFSNTID 245
              :|| |   |   : :|    :: |    | : ::: ||:  ::    || :|: || : :
Db        238 LCLKL--GPPTPLAIPNFLLSYVTRSSDNISCLIIPPLL--VQPMQFSNSSCL-FSPSYN 292

Qy        246 TTS---------SQCIRWVTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVP 295
              :|          | |   :|  |  :|  :|  |:|| : ||  |  :   :| |: |:|
Db        293 STEEIDLGHVAFSNCTS-ITNVTGPICAVNGSVFLCGNNMAYTYLPTNWTGLCVLATLLP 351

Qy        296 PMTIYTEQD-----LYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKL 350
               : |    :       :| : :|  : :  :| :   |:     ||   :  |   | ||
Db        352 DIDIIPGDEPVPIPAIDHFIYRP-KRAIQFIPLLAGLGITAAFTTGATGLGVSVTQYTKL 410

Qy        351 SQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQ 410
              | ::  |:: :: ::  ||||::||| ||||||| ||||||::|| || | |: |:|||:
Db        411 SNQLISDVQILSSTIQDLQDQVDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYVNK 470

Qy        411 SRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFN 470
              | || :|:| ::: :: | ::| :   |  |   :|::|||||||  |:|||  ||||||
Db        471 SGIVRDKIKTLQEELERRRKDLASNPLWTGLQGLLPYLLPFLGPLLTLLLLLTIGPCIFN 530

Qy        471 LLVKFVSSRIEAVKLQMVLQ 490
               |  |:: ::  :   :: |
Db        531 RLTAFINDKLNIIHAMVLTQ 550


RESULT 20
Q9WPZ9
ID   Q9WPZ9      PRELIMINARY;      PRT;   580 AA.
AC   Q9WPZ9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope glycoprotein (Fragment).
GN   Name=env;
OS   Simian retrovirus SRV-2.
OC   Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OX   NCBI_TaxID=39068;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D2/RHE/OR/V1;
RX   MEDLINE=95191045; PubMed=7884914;
RA   Marracci G.H., Kelley R.D., Pilcher K.Y., Crabtree L., Shiigi S.M.,
RA   Avery N., Leo G., Webb M.C., Hallick L.M., Axthelm M.K., Machida C.A.;
RT   "Simian AIDS type D serogroup 2 retrovirus: isolation of an infectious
RT   molecular clone and sequence analyses of its envelope glycoprotein
RT   gene and 3' long terminal repeat.";
RL   J. Virol. 69:2621-2628(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D2/RHE/OR/V1;
RX   MEDLINE=99373443;
RA   Marracci G.H., Avery N.A., Shiigi S.M., Couch G., Palmer H.,
RA   Pilcher K.Y., Nichols H., Hallick L.M., Axthelm M.K., Machida C.A.;
RT   "Molecular cloning and cell-specific growth characterization of
RT   polymorphic variants of type D serogroup 2 simian retroviruses.";
RL   Virology 261:43-58(1999).
DR   EMBL; AF126468; AAD43261.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
SQ   SEQUENCE   580 AA;  63668 MW;  84B03CEC45BE9457 CRC64;

  Query Match             21.5%;  Score 567.5;  DB 2;  Length 580;
  Best Local Similarity   28.5%;  Pred. No. 9.1e-38;
  Matches  150;  Conservative   99;  Mismatches  203;  Indels   75;  Gaps   15;

Qy         17 FALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSK---------GN---STFTA 64
              : :|||     | | | :  :      |   |:  :| :|          ||    ||
Db         54 YVITAPTVYLATVSCSSHTAY-----QPS--DSLKWRCVSNPTLANGENIGNCPCQTFKE 106

Qy         65 HTHMPRNCYNSATLCMHANTHYWTG-------------------------KMINPSCPGGLG 101
                |   :|| :   |   |  |:|                         ::: | | :|
Db        107 SVH--SSCYTTYQECFFGNKTYYTAILASNRAPTIGTSNVPTVLGNTHNLLSAGCTGNVG 164

Qy        102 ATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTH 161
                :||       :||||| | :|||  |:: : :: |       |  | | |     :
Db        165 QPICWNPKAPVHISDGGGPQDKAREIAVQKRLEEIHRSLFPELRYHPLALPKARGKEKID 224

Qy        162 TRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVG 221
               :  :|   | : |:: :      ||:|||      ::||     |:     | :  ::
Db        225 AQTFNLLTATYSLLNKSNPNLANECWLCLPSGNPVPLAIPS----NDSFLGSNLSCPIIP 280

Qy        222 P-LVSNLEITHTSNLTCV--KFSNT--------IDTTSSQCIRWVTPPTRIVCLPSGIFF 270
              | ||  ||  :  | :|:   | |         :: |:   |  :   :  :| |:   |
Db        281 PLLVQPLEFINLINASCLYSPFQNNSFGVDVGLVEFTNCSTILNI---SHSLCAPNSSVF 337

Qy        271 VCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP------HNKR--VPIL 321
              ||| : ||  |  :    | |: |:| : |       :  || |      |  |  |  :
Db        338 VCGNNKAYTYLPTNWTGTCVLATLLPDIDIVPG----DAPVPVPAIDHYLHRARRAVQFI 393

Qy        322 PFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQ 381
              | ::  |:   : ||   :  |   | :||:::  |:: :: ::  ||||::||| ||||
Db        394 PLLVGLGITTAVSTGTAGLGYSITQYTRLSRQLISDVQAISSTIQDLQDQVDSLAEVVLQ 453

Qy        382 NRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLL 441
              ||| ||||||::|| || | |: |:| |:| || :|:| ::: :: | :|: :   |  |
Db        454 NRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKRLQEDLEKRRKEIIDNPFWTGL 513
```

```
Qy        442 SQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
                 :|::|| ||||  |:||: ||| ||| :: ||  :::|:: : :
Db        514 HGLLPYLLPLLGPLLCLLLLITFGPLIFNKIIAFVKQQMDAIQAKPI 560

RESULT 21
ENV_SRV2
ID   ENV_SRV2       STANDARD;      PRT;   574 AA.
AC   P51515;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Env polyprotein (Coat polyprotein) [Contains: Coat protein GP70; Coat
DE   protein GP20].
GN   Name=ENV;
OS   Simian retrovirus SRV-2.
OC   Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OX   NCBI_TaxID=39068;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87151131; PubMed=2435057;
RA   Thayer R.M., Power M.D., Bryant M.L., Gardner M.B., Barr P.J.,
RA   Luciw P.A.;
RT   "Sequence relationships of type D retroviruses which cause simian
RT   acquired immunodeficiency syndrome.";
RL   Virology 157:317-329(1987).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16605; AAA47563.1; ALT_INIT.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT   CHAIN         1    382       Coat protein GP70.
FT   CHAIN       383    574       Coat protein GP20.
FT   TRANSMEM    386    406       Potential.
FT   TRANSMEM    515    535       Potential.
FT   CARBOHYD    117    117       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    233    233       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    260    260       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    267    267       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    288    288       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    298    298       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    312    312       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    318    318       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    327    327       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    345    345       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    475    475       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   574 AA;  62952 MW;  662B1E11A437E055 CRC64;

  Query Match             21.4%;  Score 562.5;  DB 1;  Length 574;
  Best Local Similarity   27.8%;  Pred. No. 2.3e-37;
  Matches  151;  Conservative   93;  Mismatches  192;  Indels  107;  Gaps   15;

Qy         24 PCCCT-----------------TSSSPYQEFLXRTRLPGNIDAPSYRSLSK--------- 57
              || |                  :| : ||            |:  :| :|
Db         41 PCDCAGGYVNAAPTVYLAAVSCSSHTAYQP----------SDSLKWRCVSNPTLANGENI 90

Qy         58 GN---STFTAHTHMPRNCYNSATLCMHANTHYWTG-----------------------KM 91
              ||    ||    |   :|| :   |   |  |:|                         :
Db         91 GNCPCKTFKESVH--SSCYTAYQECFFGNKTYYTAILASNRAPTIGTSNVPTVLGNTHNL 148

Qy         92 INPSCPGGLGATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVL 151
              ::  | | :|  :||       :||||| | :|||  |:: : :: :       |  | |
Db        149 LSAGCTGNVGQPICWNPKAPVHISDGGGPQDKAREIAVQKRLEEIHKSLFPELRYHPLAL 208

Qy        152 SKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFST 211
               |     :   :  :|   | : |:: :      ||:|||    |   ||:    |:
Db        209 PKARGKEKIDAQTFNLLTATYSLLNKSNPNLANECWLCLP-SGNP---IPLAIPSNDSFL 264

Qy        212 EINTTSVLVGP-LVSNLEITHTSNLTC----------------VKFSNTIDTTSSQCIRW 254
                | :  :: | ||  ||  :  |  :|                |:|:|   | :
Db        265 GSNLSCPIIPPLLVQPLEFMNLINASCFYSPFQNNSFDVDVGLVEFANCSTTLNI----- 319

Qy        255 VTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP 313
                  :  :| |:   |||| : ||  |  :    | |:|: | : |       :  || |
Db        320 ----SHSLCAPNSSVFVCGNNKAYTYLPSNWTGTCVLATLLPDIDIVPG----DAPVPVP 371

Qy        314 ------HNKR--VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
                    |  |  |  :| ::  |:   : ||   :  |   | |||:::  |:: :: ::
Db        372 AIDHYLHRARRAVQFIPLLVGLGITTAVSTGTAGLGYSITQYTKLSRQLISDVQAISSTI 431

Qy        366 VTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI 425
                ||||::||| ||||||| ||||||::|| || | |: |:| |:| || :|:| ::: :
Db        432 QDLQDQVDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKRLQEDL 491

Qy        426 QCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKL 485
              : | :|: :   |  |   :|::|| ||||  |:||: ||| ||| :: ||  :|:|::
Db        492 EKRRKEIIDNPFWTGLHGLLPYLLPLLGPLFCLLLLITFGPLIFNKIITFVKQQIDAIQA 551

Qy        486 QMV 488
              : :
Db        552 KPI 554

RESULT 22
Q98654
ID   Q98654       PRELIMINARY;      PRT;   564 AA.
AC   Q98654;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein.
GN   Name=env;
OS   RD114 retrovirus.
OC   Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
OX   NCBI_TaxID=11834;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96078985; PubMed=7494248;
RA   Cosset F.L., Takeuchi Y., Battini J.L., Weiss R.A., Collins M.K.L.;
RT   "High titer packaging cells producing recombinant retroviruses
RT   resistant to human serum.";
RL   J. Virol. 69:7430-7436(1995).
DR   EMBL; X87829; CAA61093.1; -.
DR   HSSP; P03385; 1MOF.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
SQ   SEQUENCE   564 AA;  62024 MW;  35F60EE7A51E91AC CRC64;

  Query Match             21.2%;  Score 559.5;  DB 2;  Length 564;
  Best Local Similarity   28.9%;  Pred. No. 4e-37;
  Matches  134;  Conservative   91;  Mismatches  199;  Indels   39;  Gaps    8;

Qy         63 TAHTHMPRNCYNSATLCMHANTHYWTGKMI--------------NPS------CPGGLGA 102
              |    |  :||     |   |  |:|  ::              ||:      | | :
Db         93 TFQDSMHSSCYTEYRQCRRINKTYYTATLLKIRSGSLNEVQILQNPNQLLQSPCRGSINQ 152

Qy        103 TVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHT 162
               |||:      :|||||     |   |:: : |: :  :    |  | | |: : |
Db        153 PVCWSATAPIHISDGGGPLDTKRVWTVQKRLEQIHKAMTPELQYHPLALPKVRDDLSLDA 212

Qy        163 RLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGP 222
              |   : |||   |   :     :||:|| |     ::|| |    : :  :   |  : |
Db        213 RTFDILNTTFRLLQMSNFSLAQDCWLCLKLGTPTPLAIPTPSLTYSLADSLANASCQIIP 272

Qy        223 --LVSNLEITHTSNLTCVKFSNT-------IDTTSSQCIRWVTPPTRIVCLPSGIFFVCG 273
                ||  :: :::| |:    ::|       :  |:   :  |: |   :|  :|  |:||
Db        273 PLLVQPMQFSNSSCLSSPFINDTEQIDLGAVTFTNCTSVANVSSP---LCALNGSVFLCG 329

Qy        274 TS-AYHCLNGSSESMCFLSFLVPPMTIYTEQD-----LYNHVVPKPHNKRVPILPFVIRA 327
               : ||  |  :   :|  : |:| : |    :       :| : :|  : |  :| :
Db        330 NNMAYTYLPQNWTRLCVQASLLPDIDINPGDEPVPIPAIDHYIHRP-KRAVQFIPLLAGL 388

Qy        328 GVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALD 387
              |:     ||   :  |   | ||| ::  |:: :: ::  ||||::||| ||||||| ||
Db        389 GITAAFTTGATGLGVSVTQYTKLSHQLISDVQVLSGTIQDLQDQVDSLAEVVLQNRRGLD 448

Qy        388 LLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWMPW 447
              ||||::|| || | |: |:| |:| ||  |:: ::: :| | | |     |  |  ::|:
Db        449 LLTAEQGGICLALQEKCCFYANKSGIVRNKIRTLQEELQKRRESLATNPLWTGLQGFLPY 508

Qy        448 VLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLQ 490
              :|| ||||  |:|:|  |||:|: |: |:: |:  |   :: |
Db        509 LLPLLGPLLTLLLILTIGPCVFSRLMAFINDRLNVVHAMVLAQ 551

RESULT 23
ENV_AVISN
ID   ENV_AVISN      STANDARD;      PRT;   567 AA.
AC   P31796;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Env polyprotein (Coat polyprotein) [Contains: Coat protein GP73; Coat
DE   protein GP22].
GN   Name=ENV;
OS   Avian spleen necrosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX   NCBI_TaxID=11899;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Clone PPB101;
RX   MEDLINE=92219390; PubMed=1313915;
RA   Kewalramani V.N., Panganiban A.T., Emerman M.;
RT   "Spleen necrosis virus, an avian immunosuppressive retrovirus, shares
RT   a receptor with the type D simian retroviruses.";
RL   J. Virol. 66:3026-3031(1992).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M87666; -; NOT_ANNOTATED_CDS.
DR   PIR; A38212; VCFVAS.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR008965; Cellul_bind.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT   CHAIN         1    397       Coat protein GP73.
FT   CHAIN       398    567       Coat protein GP22.
FT   TRANSMEM    402    418       Potential.
FT   CARBOHYD    245    245       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    274    274       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    280    280       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    306    306       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    319    319       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    328    328       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    335    335       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    491    491       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   567 AA;  61596 MW;  204EA57C32159175 CRC64;
```

```
  Query Match            21.2%;  Score 559.5;  DB 1;  Length 567;
  Best Local Similarity  31.9%;  Pred. No. 4e-37;
  Matches  153;  Conservative   75;  Mismatches  158;  Indels   93;  Gaps   19;

Qy         66 THMPRNCYNSATLCMHANTHYWT---------------GKMINPSCPGGLGATVCWTYFT 110
              | |   ||     |      |:|                |:|  || | :|  |||
Db        121 TQMHSTCYEKTQECTLLGKTYFTAILQKTKLGSYEDGPNKLIQASCTGTVGKPVCWDPVA 180

Qy        111 HTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPS-PYKGLVLSKLHETLRTHTRLVSLFN 169
                 :|||||     ||: |:| : :: | || ||  |  | | :        :| | |
Db        181 PVYVSDGGGPTDMIREESVRERLEEIIR-HSYPSVQYHPLALPR--------SRGVDLDP 231

Qy        170 TT---LTRLHEV-SAQNP---TNCWMCLPLHFRPYISIPVPEQWN-----NFSTEI---- 213
               |   |   |:| :| ||    |||:|: |     |   :|   |     | |  :
Db        232 QTSDILEATHQVLNATNPKLAENCWLCMTL--GTPIPAAIPTNGNVTLDGNCSLSLPFGC 289

Qy        214 ----------------NTTSVLVGPLVSNLEITHTSNLTCVKFSNTIDTTSSQCIRWVTP 257
                              | | : ||         | :| | ::    :   :||    :
Db        290 NPPGSIDVSCYAGEADNRTGIPVG-------YVHFTNCTSIQ---EVTNETSQ----MGN 335

Qy        258 PTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP--- 313
               ||: | | |  |||| : ||  |      :| |: :|| ::| : ::     :| |
Db        336 LTRL-CPPPGHVFVCGNNMAYTALPNKWIGLCILASIVPDISIISGEE----PIPLPSIE 390

Qy        314 -----HNKRVPILPFVIRAGVLGRL---GTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
                   | : |  :| ::  |: |     |||:|    |   |:||| ::  |:: :: ::
Db        391 YTARRHKRAVQFIPLLVGLGISGATLAGGTGLG---VSVHTYHKLSNQLIEDVQALSGTI 447

Qy        366 VTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI 425
                ||||::||| ||||||| ||||||::|| || | |: |:| |:| || :|:::::: :
Db        448 NDLQDQIDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIRKLQEDL 507

Qy        426 QCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVK 484
                |   | :   |  |: ::|::||  ||||  ||| |  ||||   | : :  :|:|||
Db        508 IERKRALYDNPLWSGLNGFLPYLLPLLGPLFGLILFLTLGPCIMKTLTRIIHDKIQAVK 566

RESULT 24
ENV_SRV1
ID   ENV_SRV1       STANDARD;      PRT;   587 AA.
AC   P04027;
DT   23-OCT-1986 (Rel. 02, Created)
DT   23-OCT-1986 (Rel. 02, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Env polyprotein precursor (Coat polyprotein) [Contains: Coat protein
DE   GP70; Coat protein GP20].
GN   Name=ENV;
OS   Simian retrovirus SRV-1.
OC   Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OX   NCBI_TaxID=11942;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86151668; PubMed=3006247;
RA   Power M.D., Marx P.A., Bryant M.L., Gardner M.B., Barr P.J.,
RA   Luciw P.A.;
RT   "Nucleotide sequence of SRV-1, a type D simian acquired immune
RT   deficiency syndrome retrovirus.";
RL   Science 231:1567-1572(1986).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M11841; AAA47733.1; -.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
```

```
KW   Coat protein; Glycoprotein; Polyprotein; Signal; Transmembrane.
FT   SIGNAL        1     20
FT   CHAIN        21    395       Coat protein GP70.
FT   CHAIN       396    587       Coat protein GP20.
FT   TRANSMEM    399    419       Potential.
FT   TRANSMEM    528    548       Potential.
FT   CARBOHYD    120    120       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    237    237       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    266    266       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    271    271       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    277    277       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    280    280       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    295    295       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    308    308       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    322    322       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    328    328       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    340    340       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    358    358       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    488    488       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    585    585       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   587 AA;  64474 MW;  80939DD3BFB65A8D CRC64;

  Query Match             21.1%;  Score 556;  DB 1;  Length 587;
  Best Local Similarity   29.4%;  Pred. No. 8.1e-37;
  Matches  160;  Conservative   84;  Mismatches  201;  Indels   99;  Gaps   17;

Qy          19 LTAPPPCCCTT-SSSPYQEFLXRTRL-------------------PGNIDAPSYRSLSKG 58
               :::||    || | | |  :     |                   |   :: || |:
Db          49 VSSPPTNSLTTVSCSTYTAYSVTNSLKWQCVSTPTTASPTHIGSCPSQCNSQSYDSVH-- 106

Qy          59 NSTFTAHTHMPRNCYNSATLCMHANTHYWTGKMI---NPS-------------------- 95
                            |||    |   |  | |  ||   :||
Db         107 -----------ATCYNHYQQCTIGNKTYLTATMIRDKSPSSGDGNVPTILGNNQNLIIAG 155

Qy          96 CP-GGLGATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKL 154
               ||    |  |||       |||||| | : ||  | :   :| :       |  | | :
Db         156 CPENKKGQVVCWNSQPSVHMSDGGGPQDKVREIIVNKKFEELHKSLFPELSYHPLALPEA 215

Qy         155 H--ETLRTHT-----RLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWN 207
                  | :  ||      : || | :  |      |   :||:||    |    :|:   ::
Db         216 RGKEKIDAHTFDLLATVHSLLNVSSQR------QLAEDCWLCL----RSGDPVPLALPYD 265

Qy         208 NFSTEINT-------TSVLVGP--LVSNLEITHTSNLTCVKFSNTID--------TTSSQ 250
               | |   :|        | |: |  ||     ||:  |     :|: |        |  |
Db         266 NTSCSNSTFFFNCSNCSCLITPPFLVQPFNFTHSVCLYADYQNNSFDIDVGLAGFTNCSS 325

Qy         251 CIRWVTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQD----- 304
                |  :: |:  :| |:   |||| : ||  |  :    | |: |:| : |    :
Db         326 YIN-ISKPSSPLCAPNSSVFVCGNNKAYTYLPTNWTGSCVLATLLPDIDIIPGSEPVPIP 384

Qy         305 LYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDS 364
                 :| : :|  : :  :| ||  |:   : ||   :  |   | ||| ::  |:: :: :
Db         385 AIDHFLGRP-KRAIQFIPLVIGLGITTAVSTGTAGLGVSLTQYTKLSHQLISDVQAISST 443

Qy         365 LVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDR 424
               :  ||||::||| ||||||| ||||||::|| || | |: |:| |:| || :|:| ::|
Db         444 IQDLQDQVDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDD 503

Qy         425 IQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVK 484
               :: | ::| :   |      :|:|:| ||||  |:|:| ||| ||| |: |:  :||:::
Db         504 LEKRRKQLIDNPFWTGFHGLLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQ 563

Qy         485 LQMV 488
                : :
Db         564 AKPI 567


RESULT 25
Q9WQ03
ID   Q9WQ03      PRELIMINARY;      PRT;   580 AA.
AC   Q9WQ03;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope glycoprotein (Fragment).
GN   Name=env;
OS   Simian retrovirus SRV-2.
OC   Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OX   NCBI_TaxID=39068;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D2/RHE/OR;
RX   MEDLINE=95191045; PubMed=7884914;
RA   Marracci G.H., Kelley R.D., Pilcher K.Y., Crabtree L., Shiigi S.M.,
RA   Avery N., Leo G., Webb M.C., Hallick L.M., Axthelm M.K., Machida C.A.;
RT   "Simian AIDS type D serogroup 2 retrovirus: isolation of an infectious
RT   molecular clone and sequence analyses of its envelope glycoprotein
RT   gene and 3' long terminal repeat.";
RL   J. Virol. 69:2621-2628(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D2/RHE/OR;
RX   MEDLINE=99373443;
RA   Marracci G.H., Avery N.A., Shiigi S.M., Couch G., Palmer H.,
RA   Pilcher K.Y., Nichols H., Hallick L.M., Axthelm M.K., Machida C.A.;
RT   "Molecular cloning and cell-specific growth characterization of
RT   polymorphic variants of type D serogroup 2 simian retroviruses.";
RL   Virology 261:43-58(1999).
DR   EMBL; AF126467; AAD43257.1; -.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
FT   NON_TER       1      1
SQ   SEQUENCE   580 AA;  63802 MW;  A163B6CDEAF3466A CRC64;

  Query Match             21.1%;  Score 554.5;  DB 2;  Length 580;
  Best Local Similarity   28.0%;  Pred. No. 1.1e-36;
  Matches  149;  Conservative   96;  Mismatches  201;  Indels   87;  Gaps   14;

Qy          17 FALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSK---------GN---STFTA 64
               : :|||     | | | :  :      |   |:  :| :|          ||    ||
Db          54 YVITAPTVYLATVSCSSHTAY-----QPS--DSLKWRCVSNPTLANGENIGNCPCQTFKE 106

Qy          65 HTHMPRNCYNSATLCMHANTHYWTG-----------------------KMINPSCPGGLG 101
                 |   :|| :   |   |  |:|                        :::   | | :|
Db         107 SVH--SSCYTTYQECFFGNKTYYTAILASNRAPTIGTSNVPTVLGNTHNLLSAGCTGTVG 164

Qy         102 ATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTH 161
                 :||       :||||| | :|||  |:: : :: |       |  | | |     :
Db         165 QHICWNPKAPVHISDGGGPQDKAREIAVQKRLEEIHRSLFPELRYHPLALPKARGKEKID 224

Qy         162 TRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVG 221
                :  :|   | : |:: :      ||:|||      ::||     |:     | :  ::
Db         225 AQTFNLLTATYSLLNKSNPNLANECWLCLPSGNPVPLAIPS----NDSFLGSNLSCPIIP 280

Qy         222 P-LVSNLEITHTSNLTC----------------VKFSNTIDTTSSQCIRWVTPPTRIVCL 264
               | ||  ||  :  | :|                |:|:|   | :          :  :|
Db         281 PLLVQPLEFINLINASCLYSPSQNNSFDVDVGLVEFTNCSTTLNI---------SHSLCA 331

Qy         265 PSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP------HNKR 317
               |:   |||| : ||  |  :    | |: |:| : |       :  || |      |  |
Db         332 PNSSVFVCGNNKAYTYLPTNWTGTCVLATLLPDIDIVPG----DAPVPVPAIDHYLHRAR 387

Qy         318 --VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSL 375
                 |  :| ::  |:   : ||   :  |   | |||:::  |:: :: ::  ||||::||
Db         388 RAVQFIPLLVGLGITTAVSTGTTGLGYSITQYTKLSRQLISDVQAISSTIQDLQDQVDSL 447

Qy         376 AAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNT 435
               | ||||||| ||| ||::|| || | |: |:| |:| || :|:| ::: :: | :|: :
Db         448 AEVVLQNRRGLDLFTAEQGGICLALQEKCCFYANKSGIVRDKIKRLQEDLEKRRKEIIDN 507

Qy         436 ERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
                 |  |   :|::|| | ||  |:||: ||| ||| :: ||  :::|:: : :
```

```
Db        508 PFWTGLHGLLPYLLPLLRPLLCLLLLITFGPLIFNKIIAFVKQQMDAIQAKPI 560

RESULT 26
ENV_SRV2R
ID   ENV_SRV2R      STANDARD;      PRT;   574 AA.
AC   P51520;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Env polyprotein (Coat polyprotein) [Contains: Coat protein GP70; Coat
DE   protein GP20].
GN   Name=ENV;
OS   Simian retrovirus SRV-2 (isolate 2R-18B1).
OC   Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OX   NCBI_TaxID=73490;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95191045; PubMed=7884914;
RA   Marracci G.H., Kelley R.D., Pilcher K.Y., Crabtree L., Shiigi S.M.,
RA   Avery N., Leo G., Webb M.C., Hallick L.M., Axthelm M.K., Machida A.;
RT   "Simian AIDS type D serogroup 2 retrovirus: isolation of an infectious
RT   molecular clone and sequence analyses of its envelope glycoprotein
RT   gene and 3' long terminal repeat.";
RL   J. Virol. 69:2621-2628(1995).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L38695; AAA68002.1; -.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT   CHAIN         1    382       Coat protein GP70.
FT   CHAIN       383    574       Coat protein GP20.
FT   TRANSMEM    386    406       Potential.
FT   TRANSMEM    515    535       Potential.
FT   CARBOHYD    117    117       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    233    233       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    260    260       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    267    267       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    288    288       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    298    298       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    312    312       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    318    318       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    327    327       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    345    345       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    475    475       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   574 AA;  63014 MW;  484EAF73ECABD0D6 CRC64;

  Query Match             21.0%;  Score 553.5;  DB 1;  Length 574;
  Best Local Similarity   28.0%;  Pred. No. 1.3e-36;
  Matches  149;  Conservative   96;  Mismatches  201;  Indels   87;  Gaps   14;

Qy          17 FALTAPPPCCCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSK---------GN---STFTA 64
               : :|||     | | | :  :      |   |:  :| :|          ||    ||
Db          48 YVITAPTVYLATVSCSSHTAY-----QPS--DSLKWRCVSNPTLANGENIGNCPCQTFKE 100

Qy          65 HTHMPRNCYNSATLCMHANTHYWTG-----------------------KMINPSCPGGLG 101
                 |   :|| :   |   |  |:|                         :::  | | :|
Db         101 SVH--SSCYTTYQECFFGNKTYYTAILASNRAPTIGTSNVPTVLGNTHNLLSAGCTGTVG 158

Qy         102 ATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTH 161
                 :||       :||||| | :|||  |:: : :: |       |  |  | |     :
Db         159 QHICWNPKAPVHISDGGGPQDKAREIAVQKRLEEIHRSLFPELRYHPLALPKARGKEKID 218

Qy         162 TRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVG 221
                :  :|   | : |:: :      ||:|||      ::||     |:     | :  ::
Db         219 AQTFNLLTATYSLLNKSNPNLANECWLCLPSGNPVPLAIPS----NDSFLGSNLSCPIIP 274

Qy         222 P-LVSNLEITHTSNLTC----------------VKFSNTIDTTSSQCIRWVTPPTRIVCL 264
               | ||  ||  :  | :|                |:|:|   | :          :  :|
Db         275 PLLVQPLEFINLINASCLYSPSQNNSFDVDVGLVEFTNCSTTLNI---------SHSLCA 325

Qy         265 PSGIFFVCGTS-AYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP------HNKR 317
               |:   |||| : ||  |  :    | |: |:| : |       :  || |      |  |
Db         326 PNSSVFVCGNNKAYTYLPTNWTGTCVLATLLPDIDIVPG----DAPVPVPAIDHYLHRAR 381

Qy         318 --VPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSL 375
                 |  :| ::   |:   : ||   :  |   | |||:::  |:: :: ::  ||||::||
Db         382 RAVQFIPLLVGLGITTAVSTGTTGLGYSITQYTKLSRQLISDVQAISSTIQDLQDQVDSL 441

Qy         376 AAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNT 435
               | ||||||| ||| ||::|| || | |: |:| |:| || :|:| ::: :: | :|: :
Db         442 AEVVLQNRRGLDLFTAEQGGICLALQEKCCFYANKSGIVRDKIKALQEDLEKRRKEIIDN 501

Qy         436 ERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV 488
                 |  |   :|::|| | ||  |:||: ||| ||| :: ||  :::|:: : :
Db         502 PFWTGLHGLLPYLLPLLRPLLCLLLLITFGPLIFNKIIAFVKQQMDAIQAKPI 554

RESULT 27
ENV_MPMV
ID   ENV_MPMV      STANDARD;      PRT;   586 AA.
AC   P07575;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Env polyprotein precursor (Coat polyprotein) [Contains: Coat protein
DE   GP70; Coat protein GP20].
GN   Name=ENV;
OS   Simian Mason-Pfizer virus (MPMV).
OC   Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OX   NCBI_TaxID=11855;
RN   [1]
RP   SEQUENCE FROM N.A. (CLONE 6A).
RX   MEDLINE=86189951; PubMed=2421920;
RA   Sonigo P., Barker C., Hunter E., Wain-Hobson S.;
RT   "Nucleotide sequence of Mason-Pfizer monkey virus: an
RT   immunosuppressive D-type retrovirus.";
RL   Cell 45:375-385(1986).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M12349; AAA47712.1; -.
DR   PIR; D25839; VCLJMP.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Coat protein; Glycoprotein; Polyprotein; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21    394       Coat protein GP70.
FT   CHAIN       395    586       Coat protein GP20.
FT   CARBOHYD    120    120       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    237    237       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    264    264       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    276    276       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    291    291       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    304    304       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    318    318       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    324    324       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    339    339       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    357    357       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    487    487       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   586 AA;  63882 MW;  3CE7A399D9E2F450 CRC64;

  Query Match             20.7%;  Score 546;  DB 1;  Length 586;
  Best Local Similarity   28.6%;  Pred. No. 5.4e-36;
  Matches  145;  Conservative   91;  Mismatches  209;  Indels   62;  Gaps   13;

Qy          26 CCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNSTFTAHTHMPRNCYNSATLCMHANTH 85
               | :| ::|    :     ||  :  || |:               :|||    |   |
Db          78 CVSTPTTPSNTHIG--SCPGECNTISYDSVH-------------ASCYNHYQQCNIGNKT 122

Qy          86 YWTG-----------------------KMINPSCPGG-LGATVCWTYFTHTSMSDGGGIQ 121
               | |                         :|   || |  |  |||       :||||| |
Db         123 YLTATITGDRTPAIGDGNVPTVLGTSHNLITAGCPNGKKGQVVCWNSRPSVHISDGGGPQ 182

Qy         122 GQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH--ETLRTHTRLVSLFNTTLTRLHEVS 179
                :||:  | :   :| |       |  | | :    | :  ||  : |  |  : |:
Db         183 DKARDIIVNKKFEELHRSLFPELSYHPLALPEARGKEKIDAHT--LDLLATVHSLLNASQ 240

Qy         180 AQNPTNCWMCLPLHFRPYISIPVPEQ-WNNFSTEINTTSVLVGP-LVSNLEITHTSNLTC 237
                    :||:||       :::|  :   :||:   | :  |  | ||     | :: |
Db         241 PSLAEDCWLCLQSGDPVPLALPYNDTLCSNFACLSNHSCPLTPPFLVQPFNFTDSNCLYA 300

Qy         238 VKFSNTID-----TTSSQC-----IRWVTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSES 286
                  :|: |      : : |     :   : |:  :| |:   |||| : ||  |  :
Db         301 HYQNNSFDIDVGLASFTNCSSYYNVSTASKPSNSLCAPNSSVFVCGNNKAYTYLPTNWTG 360

Qy         287 MCFLSFLVPPMTIYTEQD-----LYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSIT 341
                | |: |:| : |    :       :| : |   : : ::|  :  |:   : ||   :
Db         361 SCVLATLLPDIDIIPGSEPVPIPAIDHFLGKA-KRAIQLIPLFVGLGITTAVSTGAAGLG 419

Qy         342 TSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLG 401
                |   | ||| ::  |:: :: ::  ||||::||| ||||||| ||||||::|| || |
Db         420 VSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAEVVLQNRRGLDLLTAEQGGICLALQ 479

Qy         402 EERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILL 461
               |: |:| |:| || :|:| ::| :: |  :| :   |     ::|:|:| ||||  |:|:
Db         480 EKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPFWTSFHGFLPYVMPLLGPLLCLLLV 539

Qy         462 LLFGPCIFNLLVKFVSSRIEAVKLQMV 488
               | ||| ||| |: |:  :||::: : :
Db         540 LSFGPIIFNKLMTFIKHQIESIQAKPI 566


RESULT 28
AAC82575
ID   AAC82575     PRELIMINARY;      PRT;   586 AA.
AC   AAC82575;
DT   02-MAR-2004 (TrEMBLrel. 27, Created)
DT   02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT   02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE   Gp70 SU.
GN   ENV.
OS   Simian Mason-Pfizer virus (MPMV).
OC   Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OX   NCBI_TaxID=11855;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chappey C.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Petropoulos C.J.;
RT   "Appendix 2: Retroviral taxonomy, protein structure, sequences, and
RT   genetic maps.";
RL   (In) Coffin J.M. (eds.);
RL   RETROVIRUSES, pp.757-0, Cold Spring Harbor Laboratory Press,
RL   Cold Spring Harbor, New York, NY, USA (1997).
DR   EMBL; AF033815; AAC82575.1; -.
FT   CHAIN         1    586       gp70 SU.
FT   CHAIN       395    571       gp20 TM.
SQ   SEQUENCE   586 AA;  63882 MW;  3CE7A399D9E2F450 CRC64;

  Query Match             20.7%;  Score 546;  DB 2;  Length 586;
  Best Local Similarity   28.6%;  Pred. No. 5.4e-36;
  Matches  145;  Conservative   91;  Mismatches  209;  Indels   62;  Gaps   13;

Qy          26 CCTTSSSPYQEFLXRTRLPGNIDAPSYRSLSKGNSTFTAHTHMPRNCYNSATLCMHANTH 85
               | :| ::|    :     ||  :  || |:               :|||    |   |
Db          78 CVSTPTTPSNTHIG--SCPGECNTISYDSVH-------------ASCYNHYQQCNIGNKT 122

Qy          86 YWTG-----------------------KMINPSCPGG-LGATVCWTYFTHTSMSDGGGIQ 121
               | |                         :|   || |  |  |||       :||||| |
Db         123 YLTATITGDRTPAIGDGNVPTVLGTSHNLITAGCPNGKKGQVVCWNSRPSVHISDGGGPQ 182

Qy         122 GQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH--ETLRTHTRLVSLFNTTLTRLHEVS 179
                :||:  | :   :| |       |  | | :    | :  ||  : |  |  : |:
Db         183 DKARDIIVNKKFEELHRSLFPELSYHPLALPEARGKEKIDAHT--LDLLATVHSLLNASQ 240

Qy         180 AQNPTNCWMCLPLHFRPYISIPVPEQ-WNNFSTEINTTSVLVGP-LVSNLEITHTSNLTC 237
                    :||:||       :::|  :   :||:   | :  |  | ||     | :: |
Db         241 PSLAEDCWLCLQSGDPVPLALPYNDTLCSNFACLSNHSCPLTPPFLVQPFNFTDSNCLYA 300

Qy         238 VKFSNTID-----TTSSQC-----IRWVTPPTRIVCLPSGIFFVCGTS-AYHCLNGSSES 286
                  :|: |      : : |     :   : |:  :| |:   |||| : ||  |  :
Db         301 HYQNNSFDIDVGLASFTNCSSYYNVSTASKPSNSLCAPNSSVFVCGNNKAYTYLPTNWTG 360

Qy         287 MCFLSFLVPPMTIYTEQD-----LYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSIT 341
                | |: |:| : |    :       :| : |   : : ::|  :  |:   : ||   :
Db         361 SCVLATLLPDIDIIPGSEPVPIPAIDHFLGKA-KRAIQLIPLFVGLGITTAVSTGAAGLG 419

Qy         342 TSTQFYYKLSQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLG 401
                |   | ||| ::  |:: :: ::  ||||::||| ||||||| ||||||::|| || |
Db         420 VSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAEVVLQNRRGLDLLTAEQGGICLALQ 479

Qy         402 EERCYYVNQSRIVTEKVKEIRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILL 461
               |: |:| |:| || :|:| ::| :: |  :| :   |     ::|:|:| ||||  |:|:
Db         480 EKCCFYANKSGIVRDKIKNLQDDLERRRRQLIDNPFWTSFHGFLPYVMPLLGPLLCLLLV 539

Qy         462 LLFGPCIFNLLVKFVSSRIEAVKLQMV 488
               | ||| ||| |: |:  :||::: : :
Db         540 LSFGPIIFNKLMTFIKHQIESIQAKPI 566


RESULT 29
Q8BI41
ID   Q8BI41       PRELIMINARY;      PRT;   618 AA.
AC   Q8BI41;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched
DE   library, clone:D930020E02 product:weakly similar to ENVELOPE
DE   PROTEIN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=99279253; PubMed=10349636;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=21085660; PubMed=11217851;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
```

```
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=20499374; PubMed=11042159;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=20530913; PubMed=11076861;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK086309; BAC39647.1; -.
DR   HSSP; P03385; 1MOF.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
SQ   SEQUENCE   618 AA;  69513 MW;  1D4B54588727FC31 CRC64;

  Query Match             20.3%;  Score 534.5;  DB 2;  Length 618;
  Best Local Similarity   33.2%;  Pred. No. 5.1e-35;
  Matches  144;  Conservative   71;  Mismatches  134;  Indels   85;  Gaps   18;

Qy          98 GGLGATVCWTYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKG---LVLSK- 153
               | | |::|  |  | | |        :     |:|: :       |  : |   |: ||
Db         183 GNLSASLC-NYTMHISPSTSHENLDLSTTHTFKQAMKRPDAKWKNPLRFSGPPSLIFSKP 241

Qy         154 -----------LHETLRT------HTRLVSLFNTTL----TRLHEVS-AQNPTNCWMCL- 190
                           | :  |           : :| ||       | |: : |||:  : |
Db         242 AYYPCPTDIKHCHTSPATPWMHCPQAPFGTCYNLTLFEPDNSTHPVTMSVNPTHFKVKLQ 301

Qy         191 ------PL-HFRPYISIPVPEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCVKFSNT 243
                     || |::|     :  |:: :  ||         :  | :|  |||:    ||:
Db         302 GHRDPYPLSHYQPLTGAALSGQYSVWENEIT--------VQENWDI--TSNI----FSHL 347

Qy         244 IDTTSSQCIRWVTPPTRIVCLPSGIFFVCGTSAYHCLNGSSESMCFLSFLVPPMTIYTEQ 303
               :  : : |:             ||:||:|||| | ||  :   :| | |  | : :
Db         348 LSFSYAFCLN-----------SSGVFFLCGTSTYICLPANWSGVCTLVFQYPDIELLPN- 395

Qy         304 DLYNHVVPKP------------HNKRVP-ILPFVIRAGVLGRLGTGIGSITTSTQFYYKL 350
                  |  || |              || | : ||:   |:   |||||  : ||| :: :|
Db         396 ---NQTVPVPLFASVLSSDSVLRPKRSPHLFPFLAGLGISSALGTGIAGLATSTLYFQQL 452

Qy         351 SQEINGDMEQVTDSLVTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQ 410
               |: ::  :|::  |: |||:|::||| |||||||||||:||::||||||| || |:||||
Db         453 SKVLSETLEEIAASITTLQNQIDSLAGVVLQNRRALDLITAEKGGTCLFLQEECCFYVNQ 512

Qy         411 SRIVTEKVKEIRDRIQCRAEEL-QNTERWGL---LSQWMPWVLPFLGPLAALILLLLFGP 466
               | || :  :::::    || || |::: ||    | :|:||: ||||||  |  || ||
Db         513 SGIVRDAARKLQE----RASELGQHSDSWGQWPDLGRWLPWLTPFLGPLLFLFFLLTFGS 568

Qy         467 CIFNLLVKFVSSRI 480
               |: | | :||| |:
Db         569 CLLNCLTRFVSQRL 582


RESULT 30
Q9GLF7
ID   Q9GLF7       PRELIMINARY;      PRT;   620 AA.
AC   Q9GLF7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Envelope protein.
GN   Name=env;
OS   Trichosurus vulpecula (Brush-tailed possum).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
OX   NCBI_TaxID=9337;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21105294; PubMed=11160757;
RA   Baillie G.J., Wilkins R.J.;
RT   "Endogenous type D retrovirus in a marsupial, the common brushtail
RT   possum (Trichosurus vulpecula).";
RL   J. Virol. 75:2499-2507(2001).
DR   EMBL; AF284693; AAG28161.1; -.
DR   HSSP; P03385; 1MOF.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; TLV_coat; 1.
KW   Envelope protein.
SQ   SEQUENCE   620 AA;  67973 MW;  12E1FF147B805E7F CRC64;

  Query Match             19.9%;  Score 524.5;  DB 2;  Length 620;
  Best Local Similarity   28.1%;  Pred. No. 3.4e-34;
  Matches  134;  Conservative   77;  Mismatches  169;  Indels   97;  Gaps   12;

Qy          72 CYNSATLCMHANTHY----------------W--------TGKMINPSCPG-GLGATVCW 106
               || |   |   :  |                |        | |    ||    :|  |||
Db         110 CYTSTQQCTGKSGTYLTSRQQRAYGGSTGGDWGPIQISGPTNKYAQASCDKINIGKNVCW 169

Qy         107 TYFTHTSMSDGGGIQGQAREKQVKEAISQLTRGHSTPSPYKGLVLSK---------LHET 157
               :      :|||||   | || :||| : :: |    |  |  | | :          ::
Db         170 SLHAPIHVSDGGGPTDQIREMEVKERVDEIIRAMYPPLQYHPLALPRGRGVDLDVQTYDI 229

Qy         158 LRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFSTEINTTS 217
               | :  : ::: |  |  :          :||:|: |        |:|           |
Db         230 LASTHKALNITNPDLAK----------DCWLCMTL------GTPMPLALLTHDLSFATNC 273

Qy         218 VLVGPL----------------------VSNLEITHTSNLTCVK---FSNTIDTTSSQCIRW 254
                |  |                       :| : | |   |     |: |:: |
Db         274 ALSPPFRVQPMRPLSAPCIEAPFRNSSYDLNVGHASFALCTSNHTFNYTVERT------- 326

Qy         255 VTPPTRIVCLPSGIFFVC-GTSAYHCLNGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKP 313
                     ::|   |  ||| |  |:  |  :   :|  : ::| : | :     :  || |
Db         327 ----PHLLCPTPGRAFVCGGNMAFLALPSNWTGLCVQASILPDINIISG----DQPVPLP 378

Qy         314 H--------NKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYYKLSQEINGDMEQVTDSL 365
                         : |  :| ::  || | : ||   :  :   | ||| ::  |:: :: ::
```

```
Db          379 SIDYIAGRPKRAVAFIPLLVGLGVAGAMTTGSAGLGVAIHSYAKLSNQLINDVQTLSGTI 438

Qy          366 VTLQDQLNSLAAVVLQNRRALDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKEIRDRI 425
                  ||||::||| ||||||| ||||||::|| || | |: |:| |:| :| :|:|:::: :
Db          439 HDLQDQIDSLAEVVLQNRRGLDLLTAEQGGICLALQEKCCFYANKSGMVRDKIKKLQEEL 498

Qy          426 QCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEA 482
                  | :|| |   |  |   :|::|| ||||  |:||| ||| :|| |  || |::::
Db          499 VQRRKELLNNPLWNGLHGLLPYLLPLLGPLVGLLLLLSFGPWVFNRLTTFVKSQVDS 555

Search completed: January 22, 2005, 22:11:46
Job time : 141 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 10:09:14 ; Search time 5574.86 Seconds
                                          (without alignments)
                                          11273.477 Million cell updates/sec

Title:          US-09-319-156B-12
Perfect score:  1329
Sequence:       1 tcaaaatcgaagagctttag..........gcaaggatcccccagtaaca 1329

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters:      9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                       Description
------------------------------------------------------------------------------
     1    1326   99.8    1329  6  AR344389                AR344389 Sequence
     2    1326   99.8    1329  6  AX001030                AX001030 Sequence
     3    1326   99.8    1329  6  BD136199                BD136199 Retrovira
c    4  1316.4   99.1 163803  9  AC093531                AC093531 Homo sapi
     5  1076.8   81.0  10122  9  AY101590                AY101590 Pongo pyg
     6    1067   80.3 148711  9  AC133134                AC133134 Homo sapi
     7  1064.8   80.1  10124  9  AY101591                AY101591 Pongo pyg
     8  1036.8   78.0   2694  9  AF520477S2              AF520478 Homo sapi
     9  1036.8   78.0   2694  9  AF520489S2              AF520490 Homo sapi
    10  1036.8   78.0   2694  9  AF520505S2              AF520506 Homo sapi
    11  1036.8   78.0   2694  9  AF520529S2              AF520530 Homo sapi
    12  1035.2   77.9   2694  9  AF520483S2              AF520484 Homo sapi
    13  1035.2   77.9   2694  9  AF520487S2              AF520488 Homo sapi
    14  1035.2   77.9   2694  9  AF520493S2              AF520494 Homo sapi
    15  1035.2   77.9   2694  9  AF520495S2              AF520496 Homo sapi
    16  1035.2   77.9   2694  9  AF520497S2              AF520498 Homo sapi
    17  1035.2   77.9   2694  9  AF520503S2              AF520504 Homo sapi
    18  1035.2   77.9   2694  9  AF520509S2              AF520510 Homo sapi
    19  1035.2   77.9   2694  9  AF520513S2              AF520514 Homo sapi
    20  1035.2   77.9   2694  9  AF520515S2              AF520516 Homo sapi
    21  1035.2   77.9   2694  9  AF520517S2              AF520518 Homo sapi
    22  1035.2   77.9   2694  9  AF520519S2              AF520520 Homo sapi
    23  1035.2   77.9   2694  9  AF520521S2              AF520522 Homo sapi
    24  1035.2   77.9   2694  9  AF520525S2              AF520526 Homo sapi
    25  1035.2   77.9   2694  9  AF520533S2              AF520534 Homo sapi
    26  1035.2   77.9   2694  9  AF520535S2              AF520536 Homo sapi
    27  1035.2   77.9   2694  9  AF520537S2              AF520538 Homo sapi
    28  1035.2   77.9   2694  9  AF520549S2              AF520550 Homo sapi
    29  1035.2   77.9   2694  9  AF520553S2              AF520554 Homo sapi
    30  1035.2   77.9   2694  9  AF520561S2              AF520562 Homo sapi
    31  1035.2   77.9   2694  9  AF520563S2              AF520564 Homo sapi
    32  1035.2   77.9  56093  6  AX329572                AX329572 Sequence
    33  1035.2   77.9  56093  9  HSAC000064              AC000064 Human BAC
c   34  1035.2   77.9 149194  9  AC007566                AC007566 Homo sapi
    35  1033.6   77.8   2694  9  AF520485S2              AF520486 Homo sapi
    36  1033.6   77.8   2694  9  AF520491S2              AF520492 Homo sapi
    37  1033.6   77.8   2694  9  AF520499S2              AF520500 Homo sapi
    38  1033.6   77.8   2694  9  AF520501S2              AF520502 Homo sapi
    39  1033.6   77.8   2694  9  AF520507S2              AF520508 Homo sapi
    40  1033.6   77.8   2694  9  AF520511S2              AF520512 Homo sapi
    41  1033.6   77.8   2694  9  AF520523S2              AF520524 Homo sapi
    42  1033.6   77.8   2694  9  AF520527S2              AF520528 Homo sapi
    43  1033.6   77.8   2694  9  AF520531S2              AF520532 Homo sapi
    44  1033.6   77.8   2694  9  AF520539S2              AF520540 Homo sapi
    45  1033.6   77.8   2694  9  AF520541S2              AF520542 Homo sapi

                                  ALIGNMENTS


RESULT 1
AR344389
LOCUS       AR344389                1329 bp    mRNA    linear   PAT 17-AUG-2003
DEFINITION  Sequence 108 from patent US 6582703.
ACCESSION   AR344389
VERSION     AR344389.1  GI:33740330
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1329)
  AUTHORS   Perron,H., Beseme,F., Bedin,F., Paranhos-Baccala,G.,
            Komurian-Pradel,F., Jolivet-Reynaud,C. and Mandrand,B.
  TITLE     Isolated nucleotide sequences associated with multiple sclerosis or
            rheumatoid arthritis and a process of detecting
  JOURNAL   Patent: US 6582703-A 108 24-JUN-2003;
FEATURES             Location/Qualifiers
     source          1..1329
                     /organism="unknown"
                     /mol_type="mRNA"
ORIGIN

  Query Match             99.8%;  Score 1326;  DB 6;  Length 1329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1329;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
```

```
Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260

Qy       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGAGCAAGGATCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGAGCAAGGATCCC 1320

Qy       1321 CCAGTAACA 1329
              |||||||||
Db       1321 CCAGTAACA 1329


RESULT 2
AX001030
LOCUS       AX001030                1329 bp    DNA     linear   PAT 10-MAR-2000
DEFINITION  Sequence 12 from Patent WO9902666.
ACCESSION   AX001030
VERSION     AX001030.1  GI:7241266
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1329)
  AUTHORS   Ott,C. and Bedin,F.
  TITLE     RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR
            ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR
            DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
  JOURNAL   Patent: WO 9902666-A 12 21-JAN-1999;
            BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES             Location/Qualifiers
     source          1. .1329
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             99.8%;  Score 1326;  DB 6;  Length 1329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1329;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Qy          661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720

Qy          721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780

Qy          781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840

Qy          841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTATGG 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTATGG 900

Qy          901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960

Qy          961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

Qy         1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080

Qy         1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140

Qy         1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200

Qy         1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260

Qy         1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

Qy         1321 CCAGTAACA 1329
                |||||||||
Db         1321 CCAGTAACA 1329


RESULT 3
BD136199
LOCUS       BD136199                1329 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Retroviral nucleic material and nucleotide fragments, in
            particular, associated with multiple sclerosis and/or rheumatoid
            arthritis, for diagnostic, prophylactic and therapeutic uses.
ACCESSION   BD136199
VERSION     BD136199.1  GI:23231144
KEYWORDS    JP 2002509437-A/9.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1329)
  AUTHORS   Baccala,G.P., Pradel,F.K., Bedin,F., Sodoyer,M., Ott,C., Mallet,F.,
            Perron,H. and Mandrand,B.
  TITLE     Retroviral nucleic material and nucleotide fragments, in
            particular, associated with multiple sclerosis and/or rheumatoid
            arthritis, for diagnostic, prophylactic and therapeutic uses
  JOURNAL   Patent: JP 2002509437-A 9 26-MAR-2002;
            BIO MERIEUX
COMMENT     OS   Unidentified
            PN   JP 2002509437-A/9
            PD   26-MAR-2002
            PF   07-JUL-1998 JP 1999508255
            PR   07-JUL-1997 FR      97/08816
            PI   GLAUCIA PARAHNOS BACCALA,FLORENCE KOMURIAN PRADEL,FREDERIC  PI
              BEDIN,
            PI   MIREILLE SODOYER,CATHERINE OTT,FRANCOIS MALLET,HERVE PERRON,
            PI   BERNARD MANDRAND
            PC   C12N15/48,C12Q1/70,C07K14/15,A61K31/70
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   Retroviral nucleic material and nucleotide fragments, in  CC
                     particular,
            CC   associated with multiple sclerosis and/or
            rheumatoid arthritis,
            CC          for
            CC   diagnostic, prophylactic and therapeutic uses FH    Key
                Location/Qualifiers
            FT   source          1. .1329
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1329
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             99.8%;  Score 1326;  DB 6;  Length 1329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1329;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy            1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy           61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy          121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy          181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Qy          241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy          301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360

Qy          361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Qy          421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480

Qy          481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540

Qy          541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600

Qy          601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Qy          661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720

Qy          721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780

Qy          781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840

Qy          841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900

Qy          901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960

Qy          961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

Qy         1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080

Qy         1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140

Qy         1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200

Qy         1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260

Qy         1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

Qy         1321 CCAGTAACA 1329
                |||||||||
Db         1321 CCAGTAACA 1329


RESULT 4
AC093531/c
LOCUS       AC093531               163803 bp    DNA     linear   PRI 16-NOV-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
ACCESSION   AC093531
VERSION     AC093531.2  GI:16945981
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163803)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163803)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 163803)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15383820.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.3.
FEATURES             Location/Qualifiers
     source          1..163803
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-405L7"
ORIGIN

  Query Match             99.1%;  Score 1316.4;  DB 9;  Length 163803;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 1320;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy            1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
                | ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       117604 TTAAAATTGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 117545

Qy           61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       117544 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 117485

Qy          121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       117484 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 117425

Qy          181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       117424 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 117365

Qy          241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       117364 TTTACTCCTCTTTGGACCCTGTATCTTTAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 117305

Qy          301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       117304 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 117245

Qy          361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       117244 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 117185

Qy          421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       117184 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 117125

Qy          481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       117124 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 117065

Qy          541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       117064 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 117005

Qy          601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
                | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       117004 AGCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 116945

Qy          661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       116944 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 116885

Qy          721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     116884 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGATATAAACCCAGGCA 116825

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116824 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 116765

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     116764 CACTCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 116705

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116704 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 116645

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116644 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 116585

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116584 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 116525

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116524 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 116465

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116464 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 116405

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     116404 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAACGTGAGGCTTGCCACCATTTGGGAAGTGG 116345

Qy       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     116344 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTCTGGGAGCAAGGATCCC 116285

Qy       1321 CCAGTAACA 1329
              |||||||||
Db     116284 CCAGTAACA 116276


RESULT 5
AY101590
LOCUS       AY101590              10122 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Pongo pygmaeus isolate 1 endogenous retrovirus ERV-W, ERVWE1 locus,
            allele A, complete sequence.
ACCESSION   AY101590
VERSION     AY101590.1  GI:37544415
KEYWORDS    .
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 10122)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 10122)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1..10122
                     /organism="Pongo pygmaeus"
                     /mol_type="genomic DNA"
                     /isolate="1"
                     /isolation_source="PBMC of orangutan 1"
                     /db_xref="taxon:9600"
                     /cell_type="PBMC"
                     /note="isolated by PCR; ERVWE1 locus, allele A
                     endogenous_virus: ERV-W"
     LTR             1..780
     mRNA            join(247..811,2872..9669)
                     /note="putative mRNA transcript 1; gag-pol pseudogene"
     mRNA            join(247..811,7440..9669)
                     /product="envelope glycoprotein"
                     /note="putative mRNA transcript 2"
     mRNA            join(247..811,9104..9669)
                     /note="putative mRNA transcript 3; no coding region"
     CDS             7668..9284
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAM68169.1"
                     /db_xref="GI:37544416"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMHHP
                     GNIDAPSYRSFSKGTPTFTAHTHMPRNCYNSATLCMHGNTHYWTGKMINPSCPGGLGV
                     TVCWTYFTHTGMSDGGGVQDEAREKHVKEVISQLTQVHSTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWMCLPLAFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTNTTNSQCIRWVTPPTQVVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVVSKPRNKRVPILPFVMAAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSIAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNIGPWGLFSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYHRPLDWPASP
                     RSDVNDIKGTPPEEISTAQPLLRPNSAGSS"
     LTR             9343..10122
ORIGIN

  Query Match             81.0%;  Score 1076.8;  DB 9;  Length 10122;
  Best Local Similarity   89.2%;  Pred. No. 0;
  Matches 1186;  Conservative    0;  Mismatches  130;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db       8807 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCGAAAGAGGGGGAACCTGTTTATTTTT 8866

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||  | |||||||||||||||||
Db       8867 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATTGTCACTGAGAAAGTTAAAGA 8926

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| ||||| || ||||||||||||
Db       8927 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACATTGGACCCTGGGGCCT 8986

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              | ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       8987 CTTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 9046

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||| |||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       9047 GCTACTTCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 9106

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       9107 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 9154

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               |||||||   ||||||||||| |||||||||| | | | |||||||||||||||||  |
Db       9155 GATCTACCACAGACCCCTGGACTGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 9214

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| ||||||||||||||||||||||| ||||||| | ||||||||| |||
Db       9215 AGGCACCCCTCCTGAGGAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGG 9274

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       9275 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 9334

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
```

```
                || |||||||||||||||||||||||||||||||||||  |||||||||||| ||||||
Db        9335 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCTACTAAGAATCCCTAAGCCT 9394

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        9395 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 9454

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db        9455 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 9514

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| |||||
Db        9515 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCC 9574

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||||||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        9575 TTCGAGCCAGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 9634

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
               ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db        9635 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTGTTACGG 9694

Qy         901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
               ||  ||||||||||||| ||||  ||||||||||||||||||| ||  | |||||||||
Db        9695 CTTGAGCTGAGCTTTTGCTCGCTGTCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 9754

Qy         961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
                |||||| ||||||||  |||||| |||| |||||||||||| |||||||||||||||||
Db        9755 ACTGACTCCCATCCCTCCGGATCCTGCAGGGTGTCCACTGTGTTCCTGATCCAGCGAGGT 9814

Qy        1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
                ||||||||| ||||| ||||||||||||||||||||||||||||||||  ||||||| |
Db        9815 GCCCATTGCCGCTCCCAATCAGGCTAAAGGCTTGCCATTGTTCCTGCATCACTAAGTGTC 9874

Qy        1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
                 |||| |||||||| || ||||||||| |||||||||||||| ||||||||||  ||||
Db        9875 CAGGTTCGTCCTAATTGAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCTGTGAC 9934

Qy        1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
               |||| ||||||||||||||||||||||||||| ||||||||||| ||||||||||| | |
Db        9935 CCACAGCTTCTAATAGAGCTATAACACTCACCACATGGCCCAAGGTTCCATTCCTT-GAA 9993

Qy        1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
               ||  |||||||||||||||||||||||||||  |||||||||||||||| | |||||  |
Db        9994 TCCATGAGGCCAAGAACCCCAGGTCAGAGAACATGAGGCTTGCCACCATCTTGGAAGCAG 10053

Qy        1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
               ||| |  |||| |||| ||| || ||||||||||||||||||| || ||||||||| |||
Db       10054 CCCGCCACCATCTTGGAAGCAGCTCACCACCATCTTGGGAGCTCTGTGAGCAAGGACCCC 10113

Qy        1321 CCAGTAACA 1329
               | |||||||
Db       10114 CAAGTAACA 10122


RESULT 6
AC133134
LOCUS       AC133134              148711 bp    DNA     linear   PRI 28-NOV-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-120B7, complete sequence.
ACCESSION   AC133134
VERSION     AC133134.2  GI:25815299
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148711)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148711)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 148711)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 28, 2002 this sequence version replaced gi:22758291.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.7.
FEATURES             Location/Qualifiers
     source          1..148711
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-120B7"
ORIGIN

  Query Match             80.3%;  Score 1067;  DB 9;  Length 148711;
  Best Local Similarity   88.9%;  Pred. No. 0;
  Matches 1175;  Conservative    0;  Mismatches  143;  Indels    4;  Gaps    2;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               ||||||| ||||||||||||| ||||||||| || ||||||| |||| ||||||||||||
Db       99272 TCAAAATTGAAGAGCTTTAGATTTGCTAACCTCCGAAAGAGGAGGAATCTGTTTATTTTT 99331

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |  ||||||||||||||| ||||||||||||||  ||||| | |||||||||||||||||
Db       99332 AAAGGAAGAATGCTGTTATTATGTTAATCAATCCAGAATCGTCACTGAGAAAGTTAAAGA 99391

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| || ||| |||| || |||||||| | || |||||||||   | || ||| |||||
Db       99392 AATTCGAAATCAAATACAACGTAGAGCAAAAGAGCTTCAAAACGTGGGACACTGAGGCCT 99451

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| | ||||||||||||||||||||||||||||||||||
Db       99452 CCTCAGCCAATGGATGCCCTGGATTTTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 99511

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||| | ||| ||||||||||||| ||| ||||||||||||||||||||||||||||
Db       99512 GCTACTCATTTTTAGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAAT 99571

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| ||||||||  ||||||||||||| |||||||||||| |||| |||||
Db       99572 CGAAGCTGTAAAACTACAAATCATTCTTCAAATGGAGCCCCAGATGCAGCCCATAACTAA 99631

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                |||||||| |||||||||||| |||||||||| | ||||||||||||||||||||||||
Db       99632 GATCTACCGCGGACCCCTGGACTGGCCTGCTAGCCCATGCTCTGATGTTAATGACATTGA 99691

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||| |||||||||||||||||||||||| ||||||||| |||
Db       99692 AGGCACCCCTCCAGAGGAAATTTCAACTGCACAACCCCTACTACACCCCAATTCAGCAGG 99751

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||  || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       99752 AAGCAGTTAGAGTGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 99811

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               ||| | ||||||||||||||||||||||||||||||||| ||||||||||||| ||||||
Db       99812 GGGGGTACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCT 99871

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | ||||||||||||||||||||| ||||||||||||||||||||||| ||| |||||| |
Db       99872 AGCTGGGAAGGTGACCGCATCCACCTTTAAACATGGGGCTTGCAACTGAGCACACACCTG 99931

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
```

```
              |||||||||||||| || |||||||||||| ||||||||| ||||||||||| |||||||
Db      99932 ACCAATCAGAGAGCACAGTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCC 99991

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| ||||||||||  |||||||||  ||||||
Db      99992 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAAGGATCAGGATATAAATCCAGGCA 100051

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||  ||||| |||| ||||||||||||||||||   | |   |||||||||||||
Db     100052 TTCGAGCTGGCAACGGCAATCCCCTTTGGGTCCCCTCCTTGTAT---GGAGCTCTGTTTT 100108

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              |||||||||||||||||||||||| ||||||||||||||||||||| |||||| ||| |
Db     100109 CACTCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCTGTGTTTGTTAAGA 100168

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTT-GCCACCGTCACAGACCCGC 959
              ||  ||||||||||||| | |||||||||||||||||||| ||    |||||||||| ||
Db     100169 CTTGAGCTGAGCTTTTGCTTGCCATCCACCACTGCTGTTTCGCTGTTGTCACAGACCTGC 100228

Qy        960 TGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGG 1019
                |||||||||||||||   |||||||||| |||||||||||||||||||||||| ||||
Db     100229 CACTGACTTCCATCCCTCCAGATCCAGCAGGGTGTCCACTGTGCTCCTGATCCAGGGAGG 100288

Qy       1020 TACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGC 1079
               |||||||||||||||||||||||||||||||||| ||||||||||||| | || |||||
Db     100289 CACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGTCATTGTTCCTGCACGACTGAGTGC 100348

Qy       1080 CTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGA 1139
              |||||||  ||||||| || ||||||||| |||||||||||||| ||||||||||| |||
Db     100349 CTGGGTTCATCCTAATCGAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCCGTGA 100408

Qy       1140 CCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGT 1199
              |||||  ||||||||||||| | |||||||||| |||||||||||||||||||||||||
Db     100409 CCCACAACTTCTAATAGAGCCACAACACTCACCACATGGCCCAAGATTCCATTCCTTGGA 100468

Qy       1200 ATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTG 1259
              ||| ||||||||||||||||||||| | ||||   ||||||||||||||| | |||||||
Db     100469 ATCCGTGAGGCCAAGAACCCCAGGTGAAAGAACACGAGGCTTGCCACCATCTTGGAAGTG 100528

Qy       1260 GCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCC 1319
              ||||||||||||||| | ||  ||||||||||||||| |||||| | |||||||||| ||
Db     100529 GCCCACTGCCATTTTAGAAGTTGCCCACCACCATCTTAGGAGCTCTAGGAGCAAGGACCC 100588

Qy       1320 CC 1321
              ||
Db     100589 CC 100590


RESULT 7
AY101591
LOCUS       AY101591               10124 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Pongo pygmaeus isolate 1 endogenous retrovirus ERV-W, ERVWE1 locus,
            allele B, complete sequence.
ACCESSION   AY101591
VERSION     AY101591.1  GI:37544417
KEYWORDS    .
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 10124)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 10124)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .10124
                     /organism="Pongo pygmaeus"
                     /mol_type="genomic DNA"
                     /isolate="1"
                     /isolation_source="PBMC of orangutan 1"
                     /db_xref="taxon:9600"
                     /cell_type="PBMC"
                     /note="isolated by PCR; ERVWE1 locus, allele B
                     endogenous_virus: ERV-W"
     LTR             1. .780
     mRNA            join(247. .811,2872. .9671)
                     /note="putative mRNA transcript 1; gag-pol pseudogene"
     mRNA            join(247. .811,7440. .9671)
                     /product="envelope glycoprotein"
                     /note="putative mRNA transcript 2"
     mRNA            join(247. .811,9104. .9671)
                     /note="putative mRNA transcript 3; no coding region"
     CDS             7668. .9284
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAM68170.1"
                     /db_xref="GI:37544418"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMHHP
                     GNIDAPSYRSFSKGTPTFTAHTHMPRNCYNSATLCMHGNTHYWTGKMINPSCPGGLGV
                     TVCWTYFTHTGMSDGGGVQDEAREKHVKEVISQLTQVHSTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWMCLPLAFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTNTTNSQCIRWVTPPTQVVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVVSKPRNKRVPILPFVMAAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSIAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNIGPWGLFSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYHRPLDWPASP
                     RSDVNDIKGTPPEEISTAQPLLRPNSAGSS"
     LTR             9343. .10124
ORIGIN

  Query Match             80.1%;  Score 1064.8;  DB 9;  Length 10124;
  Best Local Similarity   89.1%;  Pred. No. 2.3e-313;
  Matches 1186;  Conservative    0;  Mismatches  130;  Indels   15;  Gaps    3;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db       8807 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCGAAAGAGGGGGAACCTGTTTATTTTT 8866

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||  | |||||||||||||||||
Db       8867 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATTGTCACTGAGAAAGTTAAAGA 8926

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| ||||| || ||||||||||||
Db       8927 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACATTGGACCCTGGGGCCT 8986

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              | ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       8987 CTTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 9046

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||| |||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       9047 GCTACTTCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 9106

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       9107 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 9154

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               |||||||   ||||||||||| |||||||||| | | | |||||||||||||||||  |
Db       9155 GATCTACCACAGACCCCTGGACTGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 9214

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| ||||||||||||||||||||||| ||||||| | ||||||||| |||
Db       9215 AGGCACCCCTCCTGAGGAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGG 9274
```

```
Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       9275 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 9334

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || |||||||||||||||||||||||||||||||||||  |||||||||||| ||||||
Db       9335 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCTACTAAGAATCCCTAAGCCT 9394

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       9395 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 9454

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       9455 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 9514

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| |||||
Db       9515 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCC 9574

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGT--CCCCTCCCATTGTATGGGAGCTCTGTT 838
              ||| |||||||||| ||||||||||||||||  |||||||| ||||||||||||||||||
Db       9575 TTCGAGCCAGCAACGGCAACCCCCTTTGGGTCCCCCCTCCCTTTGTATGGGAGCTCTGTT 9634

Qy        839 TTCACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTAT 898
              ||||  |||||||||||||||||||| |||||||||||||||||||||| ||||| |||
Db       9635 TTCATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTGTTAC 9694

Qy        899 GGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCG 958
              ||||  ||||||||||||| ||||  ||||||||||||||||||| ||  | ||||||||
Db       9695 GGCTTGAGCTGAGCTTTTGCTCGCTGTCCACCACTGCTGTTTGCCGCCACCGCAGACCCG 9754

Qy        959 CTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAG 1018
              |  |||||| ||||||||  |||||| |||| |||||||||||| |||||||||||||||
Db       9755 CCACTGACTCCCATCCCTCCGGATCCTGCAGGGTGTCCACTGTGTTCCTGATCCAGCGAG 9814

Qy       1019 GTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTG 1078
              || ||||||||| ||||| ||||||||||||||||||||||||||||||||  |||||||
Db       9815 GTGCCCATTGCCGCTCCCAATCAGGCTAAAGGCTTGCCATTGTTCCTGCATCACTAAGTG 9874

Qy       1079 CCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATG 1138
               |  |||| |||||||| || ||||||||| |||||||||||||| ||||||||||  ||
Db       9875 TCCAGGTTCGTCCTAATTGAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCTGTG 9934

Qy       1139 ACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGG 1198
              |||||| ||||||||||||||||||||||||||| ||||||||||| ||||||||||| |
Db       9935 ACCCACAGCTTCTAATAGAGCTATAACACTCACCACATGGCCCAAGGTTCCATTCCTT-G 9993

Qy       1199 TATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGT 1258
               |||  |||||||||||||||||||||||||||  |||||||||||||||| | |||||
Db       9994 AATCCATGAGGCCAAGAACCCCAGGTCAGAGAACATGAGGCTTGCCACCATCTTGGAAGC 10053

Qy       1259 GGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATC 1318
               |||| |  |||| |||| ||| || ||||||||||||||||||| || ||||||||| |
Db      10054 AGCCCGCCACCATCTTGGAAGCAGCTCACCACCATCTTGGGAGCTCTGTGAGCAAGGACC 10113

Qy       1319 CCCCAGTAACA 1329
              ||| |||||||
Db      10114 CCCAAGTAACA 10124


RESULT 8
AF520477S2
LOCUS       AF520477S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 132 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520478
VERSION     AF520478.1  GI:33410924
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 132 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17561.1"
                     /db_xref="GI:33410925"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             78.0%;  Score 1036.8;  DB 9;  Length 2694;
  Best Local Similarity   90.2%;  Pred. No. 1e-304;
  Matches 1134;  Conservative    0;  Mismatches  110;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722
```

```
Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db        2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
               ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db        2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy         901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
               ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db        2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy         961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
               ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db        2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy        1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
               |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db        2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy        1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
               ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db        2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy        1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
               |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db        2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy        1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
               ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db        2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 9
AF520489S2
LOCUS       AF520489S2           2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 22 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520490
VERSION     AF520490.1  GI:33410948
KEYWORDS    .
```

```
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 22 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17567.1"
                     /db_xref="GI:33410949"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             78.0%;  Score 1036.8;  DB 9;  Length 2694;
  Best Local Similarity   90.2%;  Pred. No. 1e-304;
  Matches 1134;  Conservative    0;  Mismatches  110;  Indels   13;  Gaps    2;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662
```

```
Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                |||||||||||  |||            |||||||| ||| |||||||||||| ||||||
Db        1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db        2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
               ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db        2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy         901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
               ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db        2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy         961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
               ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db        2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy        1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
               |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db        2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy        1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
               ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db        2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy        1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
               |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db        2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy        1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
               ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db        2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 10
AF520505S2
LOCUS       AF520505S2               2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 37 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520506
VERSION     AF520506.1  GI:33410980
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 37 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17575.1"
                     /db_xref="GI:33410981"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             78.0%;  Score 1036.8;  DB 9;  Length 2694;
  Best Local Similarity   90.2%;  Pred. No. 1e-304;
  Matches 1134;  Conservative    0;  Mismatches  110;  Indels   13;  Gaps    2;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
```

```
           ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy     241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
             ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db    1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy     301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
            ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db    1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy     361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
            ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db    1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy     421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
           || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db    1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy     481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
           ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db    1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy     541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
            || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy     601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
           | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db    2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy     661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
           |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db    2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy     721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
           |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db    2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy     781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
           ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db    2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy     841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
           ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db    2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy     901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
           ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db    2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy     961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
           ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db    2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy    1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
           |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db    2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy    1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
           ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db    2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy    1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
           |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db    2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy    1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
           ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db    2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 11
```

```
AF520529S2
LOCUS       AF520529S2             2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 78 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520530
VERSION     AF520530.1  GI:33411028
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 78 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17587.1"
                     /db_xref="GI:33411029"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             78.0%;  Score 1036.8;  DB 9;  Length 2694;
  Best Local Similarity   90.2%;  Pred. No. 1e-304;
  Matches 1134;  Conservative    0;  Mismatches  110;  Indels   13;  Gaps    2;

Qy       1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
           |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db    1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy      61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
           |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db    1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy     121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
```

```
             |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db      1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy       181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
             ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy       241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db      1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy       301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db      1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy       361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db      1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy       421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
             || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db      1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy       481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
             ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db      1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy       541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy       601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
             | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db      2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy       661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
             |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db      2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy       721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
             |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db      2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy       781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
             ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db      2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTT 2250

Qy       841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
             ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db      2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy       901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
             ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db      2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy       961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
             ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db      2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy      1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
             |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db      2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy      1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
             ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db      2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy      1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
             |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db      2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy      1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
             ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db      2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 12
AF520483S2
LOCUS       AF520483S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 147 allele B, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520484
VERSION     AF520484.1  GI:33410936
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 147 allele B"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17564.1"
                     /db_xref="GI:33410937"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy         1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
             |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db      1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy        61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
             |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609
```

```
Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 13
AF520487S2
LOCUS       AF520487S2               2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 148 allele B, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520488
VERSION     AF520488.1  GI:33410944
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 148 allele B"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17566.1"
                     /db_xref="GI:33410945"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  ||||||||||||||||||||| ||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||  |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550
```

```
Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 14
AF520493S2
LOCUS       AF520493S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 24 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520494
VERSION     AF520494.1  GI:33410956
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 24 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="Asian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17569.1"
                     /db_xref="GI:33410957"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
```

```
Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490
```

```
Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 15
AF520495S2
LOCUS       AF520495S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 24 allele B, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520496
VERSION     AF520496.1  GI:33410960
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 24 allele B"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="Asian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17570.1"
                     /db_xref="GI:33410961"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
```

```
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 16
AF520497S2
LOCUS       AF520497S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 30 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520498
VERSION     AF520498.1  GI:33410964
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 30 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17571.1"
                     /db_xref="GI:33410965"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy            1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
                |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db         1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy           61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
                |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db         1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy          121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
                |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db         1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy          181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
                ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy          241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                  ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db         1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy          301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                 ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db         1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy          361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                 ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db         1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy          421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
                || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db         1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy          481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
                ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db         1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy          541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                 || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db         1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy          601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
                | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db         2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy          661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
                |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db         2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy          721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
                |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db         2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy          781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
                ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db         2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTT 2250

Qy          841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
                ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db         2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy          901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
                ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db         2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy          961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
                ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db         2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy         1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
                |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db         2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy         1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
                ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db         2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy         1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
                |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db         2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy         1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
                ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db         2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 17
AF520503S2
LOCUS       AF520503S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 34 allele B, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520504
VERSION     AF520504.1  GI:33410976
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 34 allele B"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17574.1"
                     /db_xref="GI:33410977"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| ||||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 18
AF520509S2
LOCUS       AF520509S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 45 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520510
VERSION     AF520510.1  GI:33410988
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 45 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17577.1"
                     /db_xref="GI:33410989"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |  
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | ||||||||| 
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| |||||||||||||||||||||| 
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || |||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 19
AF520513S2
LOCUS       AF520513S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 49 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520514
VERSION     AF520514.1  GI:33410996
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 49 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="Asian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17579.1"
                     /db_xref="GI:33410997"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| |||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 20
AF520515S2
LOCUS       AF520515S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 49 allele B, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520516
VERSION     AF520516.1  GI:33411000
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 49 allele B"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="Asian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
```

```
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17580.1"
                     /db_xref="GI:33411001"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | ||||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || |||||||||||||| |||| ||||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||||||||||||||||||||||||| ||| |||||||||| ||||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  ||||||||||||||||||||||| | | | ||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| ||||||||||||| |||||||||| ||||||| | |||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || |||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | ||||||||||||||| |||||| |||||||||| ||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| |||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| |||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| ||||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| ||||||||||||||||||||||| |||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | |||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  ||||  |||||| |||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  |||||| || |||||||| |||||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| |||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | ||||||||||||||||||||||||||   |||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 21
AF520517S2
LOCUS       AF520517S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 55 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520518
VERSION     AF520518.1  GI:33411004
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 55 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
```

```
    misc_feature    43. .57
                    /note="splice acceptor site"
    CDS             284. .1900
                    /note="syncytin"
                    /codon_start=1
                    /product="envelope glycoprotein"
                    /protein_id="AAQ17581.1"
                    /db_xref="GI:33411005"
                    /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                    GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                    TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                    HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                    LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                    YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                    TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                    RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                    FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                    RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
    misc_feature    1705. .1720
                    /note="splice acceptor site"
    LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |  
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | ||||||||| 
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| |||||||||||||||||||||| 
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||  |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 22
AF520519S2
LOCUS       AF520519S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 55 allele B, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520520
VERSION     AF520520.1  GI:33411008
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 55 allele B"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
```

```
                   /sex="female"
                   /cell_type="PBMC's"
                   /note="endogenous_virus: HERV-W"
     misc_feature  43. .57
                   /note="splice acceptor site"
     CDS           284. .1900
                   /note="syncytin"
                   /codon_start=1
                   /product="envelope glycoprotein"
                   /protein_id="AAQ17582.1"
                   /db_xref="GI:33411009"
                   /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                   GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                   TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                   HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                   LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                   YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                   TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                   RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                   FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                   RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature  1705. .1720
                   /note="splice acceptor site"
     LTR           1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy        1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
            |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db     1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy       61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
            |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db     1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy      121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
            |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db     1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy      181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
            ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy      241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
              ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db     1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy      301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
             ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db     1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy      361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
             ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db     1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy      421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
            || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db     1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy      481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
            ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db     1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy      541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
             || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy      601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
            | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db     2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy      661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
            |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db     2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy      721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
            |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db     2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy      781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
            ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db     2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy      841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
            ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db     2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy      901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
            ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db     2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy      961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
            ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db     2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy     1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
            |||||||||| ||||| ||  |||||||||||||||||||||||||||||||||||||||
Db     2431 ACCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy     1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
            ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db     2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy     1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
            |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db     2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy     1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
            ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db     2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 23
AF520521S2
LOCUS       AF520521S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 6 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520522
VERSION     AF520522.1  GI:33411012
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 6 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17583.1"
                     /db_xref="GI:33411013"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||  |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 24
AF520525S2
LOCUS       AF520525S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 65 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520526
VERSION     AF520526.1  GI:33411020
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
```

```
     source          1..2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 65 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /country="Brazil"
                     /note="Metis
                     endogenous_virus: HERV-W"
     misc_feature    43..57
                     /note="splice acceptor site"
     CDS             284..1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17585.1"
                     /db_xref="GI:33411021"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705..1720
                     /note="splice acceptor site"
     LTR             1959..2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
```

```
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| |||||||||| ||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||| |||| ||||| ||||||||||||||||||||||| |||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 25
AF520533S2
LOCUS       AF520533S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 80 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520534
VERSION     AF520534.1  GI:33411036
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 80 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17589.1"
                     /db_xref="GI:33411037"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
```

```
Db        1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
               ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db        2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy         901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
               ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db        2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy         961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
               ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db        2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy        1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
               |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db        2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy        1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
               ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db        2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy        1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
               |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db        2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy        1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
               ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db        2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 26
AF520535S2
LOCUS       AF520535S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 80 allele B, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520536
VERSION     AF520536.1  GI:33411040
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 80 allele B"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="Caucasian
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17590.1"
                     /db_xref="GI:33411041"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match           77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
```

```
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  ||||||||||||||||||||| ||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | ||||||| |
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCACC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 27
AF520537S2
LOCUS       AF520537S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 81 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520538
VERSION     AF520538.1  GI:33411044
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
```

```
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1..2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 81 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43..57
                     /note="splice acceptor site"
     CDS             284..1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17591.1"
                     /db_xref="GI:33411045"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705..1720
                     /note="splice acceptor site"
     LTR             1959..2694
ORIGIN

  Query Match            77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity  90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | ||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || |||||||||||||| |||| ||||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| ||||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || |||||||| |||||||||||| |||||||||| ||||||||| | |||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| |||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| |||||||||| ||||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| |||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| |||||||||||| || |  
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| ||||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| ||||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | ||||||||| 
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||| 
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  |||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| |||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | ||||||||||||||||||||||||||   ||||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 28
AF520549S2
LOCUS       AF520549S2              2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 95 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520550
VERSION     AF520550.1  GI:33411068
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1: .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 95 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="female"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17597.1"
                     /db_xref="GI:33411069"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


RESULT 29
AF520553S2
LOCUS       AF520553S2               2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 96 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION   AF520554
VERSION     AF520554.1  GI:33411076
KEYWORDS    .
```

```
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1..2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 96 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="endogenous_virus: HERV-W"
     misc_feature    43..57
                     /note="splice acceptor site"
     CDS             284..1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17599.1"
                     /db_xref="GI:33411077"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705..1720
                     /note="splice acceptor site"
     LTR             1959..2694
ORIGIN

  Query Match             77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity   90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

                |||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 30
AF520561S2
LOCUS       AF520561S2             2694 bp    DNA     linear   PRI 11-FEB-2004
DEFINITION  Homo sapiens individual 71 allele A, envelope glycoprotein gene,
            complete cds, and 3' long terminal repeat, complete sequence.
```

```
ACCESSION   AF520562
VERSION     AF520562.1  GI:33411092
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
            Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
  TITLE     The endogenous retroviral locus ERVWE1 is a bona fide gene involved
            in hominoid placental physiology
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
   PUBMED   14757826
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
            CNRS-bioMerieux, Ecole Normale Superieure de Lyon - 46 allee
            d'Italie, Lyon 69364 cedex 07, France
FEATURES             Location/Qualifiers
     source          1. .2694
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="individual 71 allele A"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22"
                     /sex="male"
                     /cell_type="PBMC's"
                     /note="African
                     endogenous_virus: HERV-W"
     misc_feature    43. .57
                     /note="splice acceptor site"
     CDS             284. .1900
                     /note="syncytin"
                     /codon_start=1
                     /product="envelope glycoprotein"
                     /protein_id="AAQ17603.1"
                     /db_xref="GI:33411093"
                     /translation="MALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLWRMQRP
                     GNIDAPSYRSLSKGTPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGV
                     TVCWTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
                     HTRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSV
                     LVGPLVSNLEITHTSNLTCVKFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFFVCGTSA
                     YRCLNGSSESMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVLGALG
                     TGIGGITTSTQFYYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRALDLLTAE
                     RGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILP
                     FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP
                     RSDVNDIKGTPPEEISAAQPLLRPNSAGSS"
     misc_feature    1705. .1720
                     /note="splice acceptor site"
     LTR             1959. .2694
ORIGIN

  Query Match            77.9%;  Score 1035.2;  DB 9;  Length 2694;
  Best Local Similarity  90.1%;  Pred. No. 3.1e-304;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1482

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1483 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1543 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1602

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1603 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1663 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1722

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1723 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1770

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       1771 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 1830

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1890

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       1891 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2010

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2011 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2130

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2131 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2190

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2191 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| |||  |
Db       2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACAG 2310

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       2311 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 2370

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       2371 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 2430

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db       2431 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 2490

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db       2491 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db       2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db       2610 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666


Search completed: January 22, 2005, 16:14:04
```

```
Job time : 5585.86 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 11:11:14 ; Search time 609.913 Seconds
                                          (without alignments)
                                          11438.492 Million cell updates/sec

Title:          US-09-319-156B-12
Perfect score:  1329
Sequence:       1 tcaaaatcgaagagctttag..........gcaaggatcccccagtaaca 1329

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters:      8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_23Sep04:*
                1:  geneseqn1980s:*
                2:  geneseqn1990s:*
                3:  geneseqn2000s:*
                4:  geneseqn2001as:*
                5:  geneseqn2001bs:*
                6:  geneseqn2002as:*
                7:  geneseqn2002bs:*
                8:  geneseqn2003as:*
                9:  geneseqn2003bs:*
                10:  geneseqn2003cs:*
                11:  geneseqn2003ds:*
                12:  geneseqn2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                   %
Result           Query
  No.    Score   Match Length DB  ID                  Description
-------------------------------------------------------------------------
    1     1326    99.8    1329  2  AAV43219            Aav43219 Multiple
    2     1326    99.8    1329  2  AAX29704            Aax29704 Clone 5M6
    3     1326    99.8    1329 12  ADG14849             Adg14849 MSRV asso
    4     1299    97.7    1329  9  ADB84403            Adb84403 MSRV-1 as
    5   1035.2    77.9   56093  6  ABL61744            Abl61744 Colon ade
    6   1029.4    77.5    9502 10  ADF59718             Adf59718 Human con
    7   1023.2    77.0   10499  3  ABN97929            Abn97929 Human ret
    8    801.6    60.3    7466  5  AAS68626            Aas68626 DNA encod
    9    778.8    58.6    1393  4  AAS31002            Aas31002 Human dia
   10    775.4    58.3    2030  3  AAA63826            Aaa63826 Nucleotid
   11    766.8    57.7    2074  6  AAD41225            Aad41225 Human EMB
   12    747.6    56.3    2046  8  ACC46747            Acc46747 Human dit
   13    731.8    55.1    2946  2  AAX77526            Aax77526 Human sec
   14    731.8    55.1    2946  3  AAZ59468            Aaz59468 Human sec
   15    731.8    55.1    2946 10  ADC38776             Adc38776 Human cDN
   16    730.8    55.0    2930  6  AAD24195            Aad24195 Human syn
   17    730.2    54.9    2781  5  AAF55630            Aaf55630 Nucleotid
   18    728.6    54.8    6394  5  AAS84210            Aas84210 DNA encod
   19    722.6    54.4    7582  2  AAX25665            Aax25665 Complete
   20    722.6    54.4    7582  3  AAA59215            Aaa59215 Human end
   21    721.2    54.3    1136  2  AAX25660            Aax25660 Human end
```

```
   22    721.2    54.3    1136  3  AAA59210            Aaa59210 3' pol ge
   23    712.6    53.6    2782  2  AAX25661            Aax25661 Human end
   24    712.6    53.6    2782  3  AAA59211            Aaa59211 5' non co
   25    712.6    53.6    2782  5  AAH20069            Aah20069 HERV-W en
   26    704.8    53.0    1894  4  AAI14608            Aai14608 Probe #45
   27    704.8    53.0    1894  4  ABA56337            Aba56337 Human foe
   28    704.8    53.0    1894  4  AAI35980            Aai35980 Probe #46
   29    704.8    53.0    1894  4  ABA45822            Aba45822 Human bre
   30    704.8    53.0    1894  4  ABA25978            Aba25978 Probe #44
   31    704.8    53.0    1894  4  AAK30018            Aak30018 Human bon
   32    704.8    53.0    1894  4  AAK04516            Aak04516 Human bra
   33    704.8    53.0    1894  4  ABS29670            Abs29670 Human liv
   34    704.8    53.0    1894  5  AAI04422            Aai04422 Probe #44
   35    704.8    53.0    1894  6  ABS04589            Abs04589 Human gen
   36    680.6    51.2    2942  5  AAS77313            Aas77313 DNA encod
   37      651    49.0     893  5  AAS65964            Aas65964 DNA encod
   38    643.4    48.4   46340  3  ABN97978            Abn97978 Human ret
   39    639.8    48.1     783  2  AAX25669            Aax25669 Human end
   40    610.4    45.9    3464 10  ADE09587             Ade09587 Novel DNA
   41      607    45.7    2784  3  ABN97930            Abn97930 Human ret
c  42    601.4    45.3  157090 12  ADO47194             Ado47194 DNA seque
   43    596.8    44.9     711  3  ABN97947            Abn97947 Human ret
c  44    593.6    44.7    2316  9  ACC84313            Acc84313 Human ear
   45    580.6    43.7     711  3  ABN97946            Abn97946 Human ret


                              ALIGNMENTS


RESULT 1
AAV43219
ID   AAV43219 standard; cDNA; 1329 BP.
XX
AC   AAV43219;
XX
DT   29-DEC-1998  (first entry)
XX
DE   Multiple sclerosis associated retrovirus fragment 6.
XX
KW   Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW   env gene; rheumatoid arthritis-associated virus; ss.
XX
OS   Multiple sclerosis associated retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..490
FT                   /*tag=  a
FT                   /product= "Encodes protein AAW71069"
FT                   /transl_except= (pos:77-79, appears to code for a stop
FT                   codon)
FT                   /transl_except= (pos:125-127, appears to code for a stop
FT                   codon)
FT                   /transl_except= (pos:137-139, appears to code for a stop
FT                   codon)
XX
PN   WO9823755-A1.
XX
PD   04-JUN-1998.
XX
PF   26-NOV-1997;   97WO-IB001482.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B;
XX
DR   WPI; 1998-322732/28.
DR   P-PSDB; AAW71069.
XX
PT   New nucleic acid from retroviruses - useful for diagnosis, prevention and
PT   treatment of, e.g. multiple sclerosis.
```

```
XX
PS   Disclosure; Page 187-188; 286pp; English.
XX
CC   The present sequence represents a multiple sclerosis (MS) associated
CC   retrovirus (MSRV) genomic fragment used in the method of the invention.
CC   The invention provides complete or partial genomic sequences of the MSRV-
CC   1 pol gene, gag gene and env gene, and polypeptides encoded by these
CC   genes. The invention also provides antibodies raised against the
CC   polypeptides. The genomic sequences, polypeptides and antibodies are also
CC   claimed useful for diagnosing infection by MS and rheumatoid arthritis-
CC   associated viruses, and also for prevention and treatment of infection
CC   with these viruses
XX
SQ   Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;

  Query Match             99.8%;  Score 1326;  DB 2;  Length 1329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1329;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTATGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTATGG 900

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260

Qy       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

Qy       1321 CCAGTAACA 1329
              |||||||||
Db       1321 CCAGTAACA 1329


RESULT 2
AAX29704
ID   AAX29704 standard; DNA; 1329 BP.
XX
AC   AAX29704;
XX
DT   17-OCT-2003  (revised)
DT   27-AUG-2003  (revised)
DT   08-JUN-1999  (first entry)
XX
DE   Clone 5M6 from MSRV-1.
XX
KW   Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KW   rheumatoid polyarthritis; ss.
XX
OS   Viruses.
XX
PN   FR2765588-A1.
XX
PD   08-JAN-1999.
XX
PF   07-JUL-1997;   97FR-00008816.
XX
PR   07-JUL-1997;   97FR-00008816.
XX
PA   (INMR ) BIO MERIEUX.
XX
DR   WPI; 1999-098275/09.
DR   P-PSDB; AAW99554.
XX
PT   Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT   multiple sclerosis or rheumatoid polyarthritis.
XX
PS   Claim 1; Page 39-40; 83pp; French.
```

```
XX
CC   This sequence represents clone 5M6 from a novel multiple sclerosis
CC   related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC   prophylactic or therapeutic compositions to inhibit expression of a
CC   multiple sclerosis related virus and/or virus associated with rheumatoid
CC   polyarthritis. (Updated on 27-AUG-2003 to correct OS field.) (Updated on
CC   17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;

  Query Match             99.8%;  Score 1326;  DB 2;  Length 1329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1329;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
```

```
Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260

Qy       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

Qy       1321 CCAGTAACA 1329
              |||||||||
Db       1321 CCAGTAACA 1329


RESULT 3
ADG14849
ID   ADG14849 standard; cDNA; 1329 BP.
XX
AC   ADG14849;
XX
DT   26-FEB-2004  (first entry)
XX
DE   MSRV associated cDNA #11.
XX
KW   ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX
OS   Multiple sclerosis associated retrovirus.
XX
PN   US2003198647-A1.
XX
PD   23-OCT-2003.
XX
PF   03-APR-2002; 2002US-00114104.
XX
PR   26-NOV-1996;   96US-00756429.
PR   26-NOV-1997;   97US-00979847.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2004-032461/03.
XX
PT   New isolated nucleic acid and their fragments having the pol gene of a
PT   retrovirus, useful for diagnosing, preventing and/or treating multiple
PT   sclerosis and/or rheumatoid arthritis.
XX
PS   Disclosure; SEQ ID NO 108; 193pp; English.
XX
CC   The invention relates to an isolated nucleic acid which comprises the pol
CC   gene of a retrovirus associated with multiple sclerosis or rheumatoid
CC   arthritis. The methods and compositions of the present invention are
CC   useful for diagnosing, preventing and/or treating multiple sclerosis
CC   and/or rheumatoid arthritis. The present sequence is used in the
CC   exemplification of the invention.
XX
SQ   Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;

  Query Match             99.8%;  Score 1326;  DB 12;  Length 1329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1329;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGAACCTGTTTATTTTT 60

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260

Qy       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

Qy       1321 CCAGTAACA 1329
              |||||||||
Db       1321 CCAGTAACA 1329


RESULT 4
ADB84403
ID   ADB84403 standard; DNA; 1329 BP.
XX
AC   ADB84403;
XX
DT   04-DEC-2003  (first entry)
XX
DE   MSRV-1 associated DNA sequence #21.
XX
KW   ds; multiple sclerosis; rheumatoid arthritis; gag; pol;
KW   reverse transcriptase; ribonuclease H.
XX
OS   Unidentified;.
XX
PN   US2003039664-A1.
XX
PD   27-FEB-2003.
XX
PF   26-NOV-1997;   97US-00979847.
XX
PR   26-NOV-1996;   96US-00756429.
XX
PA   (PERR/) PERRON H.
PA   (BESE/) BESEME F.
PA   (BEDI/) BEDIN F.
PA   (PARA/) PARANHOS-BACCALA G.
PA   (KOMU/) KOMURIAN-PRADEL F.
PA   (JOLI/) JOLIVET-REYNAUD C.
PA   (MAND/) MANDRAND B.
PA   (GARS/) GARSON J A.
PA   (TUKE/) TUKE P W.
XX
PI   Perron H,  Beseme F,  Bedin F,  Paranhos-Baccala G;
PI   Komurian-Pradel F,  Jolivet-Reynaud C,  Mandrand B,  Garson JA,  Tuke PW;
XX
DR   WPI; 2003-512253/48.
XX
PT   New isolated or purified nucleic acid associated with multiple sclerosis
PT   and/or rheumatoid arthritis, useful for detecting a virus associated with
PT   multiple sclerosis or rheumatoid arthritis in a biological sample.
XX
PS   Claim 31; Page 80; 193pp; English.
```

```
XX
CC   The invention relates to an isolated or purified nucleic acid from a
CC   virus associated with multiple sclerosis and/or rheumatoid arthritis,
CC   multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
CC   pol, gag or reverse transcriptase genes (or their fragments) encoding the
CC   proteins or defined peptides (including immunodominant peptides,
CC   antigenic peptides or conserved motifs).Also included are a process for
CC   detecting a virus associated with multiple sclerosis or rheumatoid
CC   arthritis in a biological sample, a nucleic acid probe for the detection
CC   of a virus associated with multiple sclerosis or rheumatoid arthritis, a
CC   primer for the amplification by polymerisation of a nucleic acid of a
CC   viral material associated with multiple sclerosis or rheumatoid
CC   arthritis, a polypeptide exhibiting an inhibitory activity on the
CC   proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC   and an antibody directed against the MSRV-1 virus obtained by
CC   immunologically reacting a human or animal body or cells with an
CC   immunogenic agent consisting of the antigenic polypeptide defined above.
CC   The nucleic acids are useful for detecting a biological sample a virus
CC   associated with multiple sclerosis or rheumatoid arthritis, or for
CC   detecting in a biological sample, the presence of or exposure to a virus
CC   associated with multiple sclerosis or rheumatoid arthritis. The present
CC   sequence is a claimed MSRV-associated sequence whose identity cannot be
CC   accurately determined. Note: The SEQ ID numbers for the sequences as
CC   displayed in the main body of the patent do not match the SEQ ID numbers
CC   in the sequence listing. Consequently those sequences mentioned in the
CC   claims may not be the sequences the authors intended to claim.
XX
SQ   Sequence 1329 BP; 339 A; 346 C; 282 G; 332 T; 0 U; 30 Other;

  Query Match             97.7%;  Score 1299;  DB 9;  Length 1329;
  Best Local Similarity   98.0%;  Pred. No. 0;
  Matches 1302;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGN 120

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCN 180

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATN 240

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAN 300

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAN 360

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGN 420

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGN 480

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGN 540

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCN 600

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCN 660

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCN 720

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCN 780

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTN 840

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGN 900

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCN 960

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGNN 1020

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGNN 1080

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGNN 1140

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGNN 1200

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTNN 1260

Qy       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCNN 1320

Qy       1321 CCAGTAACA 1329
              |||||||||
Db       1321 CCAGTAACA 1329


RESULT 5
ABL61744
ID   ABL61744 standard; DNA; 56093 BP.
XX
AC   ABL61744;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Colon adenocarcinoma related gene sequence SEQ ID NO:81.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
```

```
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 81; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 U; 0 Other;

  Query Match             77.9%;  Score 1035.2;  DB 6;  Length 56093;
  Best Local Similarity   90.1%;  Pred. No. 0;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db      37018 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 37077

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db      37078 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 37137

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db      37138 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 37197

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      37198 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 37257

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db      37258 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 37317

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db      37318 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 37365

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db      37366 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 37425

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db      37426 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 37485

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db      37486 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 37545

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      37546 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 37605

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db      37606 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 37665

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db      37666 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 37725

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db      37726 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 37785

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db      37786 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 37845

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db      37846 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 37905

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db      37906 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 37965
```

```
Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db      37966 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 38025

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db      38026 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 38085

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db      38086 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 38145

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db      38146 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 38204

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db      38205 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 38261


RESULT 6
ADF59718
ID   ADF59718 standard; cDNA; 9502 BP.
XX
AC   ADF59718;
XX
DT   12-FEB-2004  (first entry)
XX
DE   Human contig polynucleotide sequence SEQ ID NO:2085.
XX
KW   biological activity; genetic engineering; hybridisation probe; oligomer;
KW   primer; chromosome mapping; gene mapping; recombinant protein production;
KW   human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003080795-A2.
XX
PD   02-OCT-2003.
XX
PF   09-AUG-2002; 2002WO-US025485.
XX
PR   09-AUG-2001; 2001US-0311261P.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Yang Y,  Wang Z,  Weng G,  Ma Y;
XX
DR   WPI; 2003-876918/81.
DR   P-PSDB; ADF60170.
XX
PT   New polynucleotides, useful as hybridization probes, oligomers or
PT   primers, for chromosome or gene mapping, for the recombinant production
PT   of proteins, and for generating antisense DNA or RNA.
XX
PS   Example 2; SEQ ID NO 2085; 571pp; English.
XX
CC   The present invention describes isolated polynucleotide sequences (I),
CC   which encode polypeptides (II) with biological activity. Also described:
CC   (1) a vector comprising (I); (2) an expression vector comprising (I); (3)
CC   a host cell genetically engineered to comprise (I) which is operatively
CC   associated with a regulatory sequence that modulates expression of (I) in
CC   the host cell; (4) a polypeptide (II) encoded by (I); (5) a composition
CC   comprising the polypeptide of (4) and a carrier; (6) an antibody directed
CC   against the polypeptide of (4); (7) detecting (I) or the polypeptide of
CC   (4) in a sample; (8) identifying a compound that binds to the polypeptide
CC   of (4); (9) producing the polypeptide of (4); and (10) a collection of
CC   polynucleotides comprising at least one of the polynucleotide sequences
CC   (I). The polynucleotides (I) can be used as hybridisation probes,
CC   oligomers or primers, for chromosome or gene mapping, for the recombinant
CC   production of proteins, and for generating antisense DNA or RNA. The
```

```
CC   present sequence represents a human contig polynucleotide sequence, which
CC   is used in an example from the present invention.
XX
SQ   Sequence 9502 BP; 2813 A; 2433 C; 1992 G; 2263 T; 0 U; 1 Other;

  Query Match             77.5%;  Score 1029.4;  DB 10;  Length 9502;
  Best Local Similarity   87.9%;  Pred. No. 0;
  Matches 1178;  Conservative    0;  Mismatches  129;  Indels   33;  Gaps    4;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  ||||||||||||||||||| |||||
Db       2575 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTACTTTT 2634

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       2635 AGGGGAAGAATGCTGTTACTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2694

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       2695 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2754

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2755 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2814

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       2815 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2874

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       2875 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2922

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2923 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2982

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2983 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 3042

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       3043 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 3102

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       3103 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 3162

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||| ||| |||||| ||||||||| ||||||||||||||||||||||||||
Db       3163 AGCTGGGAAGGTAACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCG 3222

Qy        661 ACCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAA 711
              |||||||         ||||||||||||||||||||||| ||||||||||||||||||||
Db       3223 ACCAATCAGGTAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAA 3282

Qy        712 GCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAA 771
              | ||||||||||||||||||||||||||||||||||||| |||||| |||  ||||||||
Db       3283 GAAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAATGATCAGGATATAA 3342

Qy        772 ACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAG 831
              || ||||||||| |||||||||| || |||| |||||||||||||||| |||||||||||
Db       3343 ACCCAGGCATTCCAGCCAGCAACGGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAG 3402

Qy        832 CTCTGTTTTCACTCTATTTCACTCTATTAAATCATGCAACTGCA--CTCTTCTGGTCCGT 889
              ||||||          || |||||||||||||| ||||||||||  ||||||||||||||
Db       3403 CTCTGT----------TTCCACTCTATTAAATCTTGCAACTGCACTCTCTTCTGGTCCGT 3452

Qy        890 GTTTTTTATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTC 949
              |||| ||| || |  ||||||||||||| |     ||||||| || | |||||| | ||
Db       3453 GTTTGTTACGGATTGAGCTGAGCTTTTGCTTATTGTCCACCATTGTTCTTTGCCGCTGTT 3512
```

```
Qy        950 ACAGACCCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGA 1009
               |||||| ||  ||||||||||||||| ||||||| |||| |||||| ||||||||||||
Db       3513 GCAGACCTGCCACTGACTTCCATCCCTCTGGATCCTGCAGGGTGTCCGCTGTGCTCCTGA 3572

Qy       1010 TCCAGCGAGGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCAT 1069
              ||||||||||  ||||||||| ||||||||| |||||||||||||||||||||||||||
Db       3573 TCCAGCGAGGCGCCCATTGCCGCTCCCGATCGGGCTAAAGGCTTGCCATTGTTCCTGCAC 3632

Qy       1070 GGCTAAGTGCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTC 1129
              |||||||||||||||||||| ||||| || |||||||||  |||||||||||||||||||
Db       3633 GGCTAAGTGCCTGGGTTTGTTCTAATTGAGCTGAACACTAATCACTGGGTTCCATGGTTC 3692

Qy       1130 TCTTCCATGACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCC 1189
              |||||  ||||||||||||||||||||| |||||||||| ||| ||||||||||||||||
Db       3693 TCTTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACCACATGGCCCAAGATTCC 3752

Qy       1190 ATTCCTTGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCAT 1249
              ||||||||| ||| |||||||||||||| |||||||||||||   |||||||||||||||
Db       3753 ATTCCTTGGAATCCGTGAGGCCAAGAACTCCAGGTCAGAGAATACGAGGCTTGCCACCAT 3812

Qy       1250 TTGGGAAGTGGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGA 1309
               | ||||| ||||  || || | |||| || ||  ||||||||||||||||||| || ||
Db       3813 CTTGGAAGCGGCCTGCTACCGTCTTGGAAGTGGTTCACCACCATCTTGGGAGCTCTGTGA 3872

Qy       1310 GCAAGGATCCCCCAGTAACA 1329
              ||||||| ||||| ||||||
Db       3873 GCAAGGACCCCCCGGTAACA 3892


RESULT 7
ABN97929
ID   ABN97929 standard; DNA; 10499 BP.
XX
AC   ABN97929;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral sequence HERV-7q.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR001513.
XX
PR   23-JUN-1998;   98FR-00007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT   for diagnosis, treatment and prevention of autoimmune and neurological
PT   diseases.
XX
PS   Claim 3; Fig 1; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention
XX
SQ   Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 U; 0 Other;

  Query Match             77.0%;  Score 1023.2;  DB 3;  Length 10499;
  Best Local Similarity   90.1%;  Pred. No. 0;
  Matches 1132;  Conservative    0;  Mismatches  111;  Indels   14;  Gaps    3;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       9018 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 9077

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       9078 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 9137

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       9138 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 9197

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       9198 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 9257

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       9258 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 9317

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       9318 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 9365

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       9366 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 9425

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       9426 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 9485

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| | | || ||||||||||||||||| |||| ||||||||||||||||
Db       9486 AAGCAGTTAGAGC-GGTCTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 9544

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       9545 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 9604

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       9605 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 9664

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       9665 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 9724

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       9725 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 9784

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       9785 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 9844

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db       9845 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 9904

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db       9905 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 9964

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db       9965 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 10024

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db      10025 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 10084

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db      10085 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 10144

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db      10145 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 10203

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db      10204 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 10260


RESULT 8
AAS68626
ID   AAS68626 standard; cDNA; 7466 BP.
XX
AC   AAS68626;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #4430.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG04439.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 4430; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 0 U; 1 Other;

  Query Match             60.3%;  Score 801.6;  DB 5;  Length 7466;
  Best Local Similarity   88.0%;  Pred. No. 9.6e-252;
  Matches  897;  Conservative    0;  Mismatches  112;  Indels   10;  Gaps    2;

Qy        321 TAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCCCTGG 380
              ||   || |||||||| ||| |||||||||||| |||||| ||||||||  |||||||||
Db       3490 TAAAACTACAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGG 3549

Qy        381 ACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAGGAAA 440
              ||||||||| ||| | | | |||||||||||||||||  ||| ||||||||| |||||||
Db       3550 ACCGGCCTGTTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAA 3609

Qy        441 TCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGTTGTC 500
              ||||| |||||||||| ||||||| | ||||||||| |||||||||||||||| || |||
Db       3610 TCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTC 3669

Qy        501 AGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGACAGGA 560
               |||||||||||||||||| |||| ||||||||||||||| || ||||||||||||||||
Db       3670 GGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGACAGGA 3729

Qy        561 CTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACCGCAT 620
              ||||||||||||||||||||||| ||||||||| ||||||| |||||||||||||| |||
Db       3730 CTAGCTGGATTTCCTAGGCTGATTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACAT 3789

Qy        621 CCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTA 680
              ||| ||||||||| |||||||||||||||||||||||| |||||||||||||||||||||
Db       3790 CCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTA 3849

Qy        681 AAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTGAGAG 740
              |||||||||| ||||||| ||||||||||||| ||||||||||||| |||||||||||||
Db       3850 AAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATTTATTGCCTGAGAG 3909

Qy        741 CACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAGCA-A 799
              |||||| ||| |||||| ||| ||||||||||| ||||||||| |||  |||||||||
Db       3910 CACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCATTCGAGCTGGCAACAGCAGC 3969

Qy        800 CCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCTATT 859
              ||||||||||||||| |||| ||||||||||||||||||||||  |||||||||||||||
Db       3970 CCCCCTTTGGGTCCCTTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCTATT 4029

Qy        860 AAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCTGAGCTTTTGTT 919
              ||||| |||||||||||||||||||||| ||||| ||| ||||| ||||||||||||| |
Db       4030 AAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGGCTCGAGCTGAGCTTTTGCT 4089

Qy        920 CGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTTCCATCCCTTTG 979
              | || ||||||||||||||||||||||  | |||||| || ||||||| |||||||| ||
Db       4090 CACCGTCCACCACTGCTGTTTGCCACCACCGCAGACCTGCCGCTGACTCCCATCCCTCTG 4149

Qy        980 GATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCATTGCCACTCCCGAT 1039
              ||||| |||| |||||| |||||||||||||||||||| |  ||||||||| ||||| ||
Db       4150 GATCCTGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAAGCGCCCATTGCCGCTCCCAAT 4209

Qy       1040 CAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTCCTAATAGAA 1099
                ||||||||||||||||||||||||||| |||||||||||||||||||| ||||| ||
Db       4210 TGGGCTAAAGGCTTGCCATTGTTCCTGCACGGCTAAGTGCCTGGGTTTGTTCTAATTGAG 4269

Qy       1100 CTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCTTCTAAT----- 1154
              ||||||||| |||||||||||||||||||||||||  ||||||||||||||||||
Db       4270 CTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACGGCTTCTAATAGAGC 4329

Qy       1155 ----AGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAGGC 1210
```

```
                 |||||||||||||||||||||||||||||| |||| |||||||  ||| |||||||
Db      4330 TAACAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCGTTCCTTGTAATCCGTGAGGC 4389

Qy      1211 CAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGCCA 1270
             |||||||||||||||||||||    |||||||||||||| | ||||| ||||  || ||
Db      4390 CAAGAACCCCAGGTCAGAGAACAAAAGGCTTGCCACCATCTTGGAAGCGGCCTGCTACCG 4449

Qy      1271 TTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCCCCAGTAACA 1329
             | |||| || ||  ||||||||||||||||||| || ||||||||| ||||| ||||||
Db      4450 TCTTGGAAGTGGTTCACCACCATCTTGGGAGCTCTGTGAGCAAGGACCCCCCGGTAACA 4508


RESULT 9
AAS31002
ID   AAS31002 standard; cDNA; 1393 BP.
XX
AC   AAS31002;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Human diagnostic and therapeutic polynucleotide (DITHP) #17.
XX
KW   Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW   cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW   acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW   respiratory disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200162927-A2.
XX
PD   30-AUG-2001.
XX
PF   21-FEB-2001; 2001WO-US006059.
XX
PR   24-FEB-2000; 2000US-0184693P.
PR   24-FEB-2000; 2000US-0184697P.
PR   24-FEB-2000; 2000US-0184698P.
PR   24-FEB-2000; 2000US-0184768P.
PR   24-FEB-2000; 2000US-0184769P.
PR   24-FEB-2000; 2000US-0184770P.
PR   24-FEB-2000; 2000US-0184771P.
PR   24-FEB-2000; 2000US-0184772P.
PR   24-FEB-2000; 2000US-0184773P.
PR   24-FEB-2000; 2000US-0184774P.
PR   24-FEB-2000; 2000US-0184776P.
PR   24-FEB-2000; 2000US-0184777P.
PR   24-FEB-2000; 2000US-0184797P.
PR   24-FEB-2000; 2000US-0184813P.
PR   24-FEB-2000; 2000US-0184837P.
PR   24-FEB-2000; 2000US-0184841P.
PR   24-FEB-2000; 2000US-0185213P.
PR   24-FEB-2000; 2000US-0185216P.
PR   12-MAY-2000; 2000US-0203785P.
PR   15-MAY-2000; 2000US-0204226P.
PR   16-MAY-2000; 2000US-0204525P.
PR   16-MAY-2000; 2000US-0204821P.
PR   16-MAY-2000; 2000US-0204908P.
PR   16-MAY-2000; 2000US-0205232P.
PR   17-MAY-2000; 2000US-0204815P.
PR   17-MAY-2000; 2000US-0204863P.
PR   17-MAY-2000; 2000US-0205221P.
PR   17-MAY-2000; 2000US-0205285P.
PR   17-MAY-2000; 2000US-0205286P.
PR   17-MAY-2000; 2000US-0205287P.
PR   17-MAY-2000; 2000US-0205323P.
PR   17-MAY-2000; 2000US-0205324P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Panzer SR,  Spiro PA,  Banville SC,  Shah P,  Chalup MS,  Chang SC;
PI   Chen A,  D'sa SA,  Amshey S,  Dahl CR,  Dam TC,  Daniels SE,  Dufour GE;
PI   Flores V,  Fong WT,  Greenawalt LB,  Hillman JL,  Jones AL,  Liu TF;
PI   Roseberry AM,  Rosen BH,  Russo FD,  Stockdreher TK,  Daffo A;
PI   Wright RJ,  Yap PE,  Yu JY,  Bradley DL,  Bratcher SR,  Chen W;
PI   Cohen HJ,  Hodgson DM,  Lincoln SE,  Jackson S;
XX
DR   WPI; 2001-502867/55.
DR   P-PSDB; AAU19431.
XX
PT   Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT   enzymes, hormones and receptors, useful in diagnostics and therapeutics.
XX
PS   Claim 1; Page 304; 522pp; English.
XX
CC   The invention relates to polynucleotides (I) encoding diagnostic and
CC   therapeutic (DITHP) polypeptides (II), which include e.g. enzymes, and
CC   proteins involved in growth and development and receptors. (I) and (II)
CC   may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate DITHP expression. For example, (I) and (II)
CC   may be used to treat disorders associated with decreased polypeptide
CC   expression by rectifying mutations or deletions in a patient's genome,
CC   that affect the activity of the DITHPs, by expressing inactive proteins
CC   or supplementing the patient's own production of them. (I) and (II) may
CC   be used to treat diseases, for example, cell proliferative disorder,
CC   Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC   leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC   (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the protein. (I) and
CC   its complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar nucleic acids in
CC   samples, and therefore which patients may be in need of restorative
CC   therapy. (II) may also be used as antigens in the production of
CC   antibodies against DITHPs and in assays to identify modulators of DITHP
CC   expression and activity. The anti-DITHP antibodies and antagonists may
CC   also be used to down regulate expression and activity. The anti-DITHP
CC   antibodies may also be used as diagnostic agents for detecting the
CC   presence of DITHPs in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA)). AAS30986-AAS31196 represent human diagnostic and therapeutic
CC   (DITHP) polynucleotides of the invention
XX
SQ   Sequence 1393 BP; 375 A; 352 C; 304 G; 360 T; 0 U; 2 Other;

  Query Match             58.6%;  Score 778.8;  DB 4;  Length 1393;
  Best Local Similarity   87.2%;  Pred. No. 1.1e-244;
  Matches  904;  Conservative    0;  Mismatches  112;  Indels   21;  Gaps    4;

Qy       302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
             ||||||||||| |||| ||| |||||||||| ||| |||||||||||||||||||||||
Db         1 GAAGCTGTAAAACTACTAATTGTTCTTCAAAAGGAGCCCCAGATGCAGTCCATGACTAAG 60

Qy       362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
             |||||||   |||||||||||||||||||||| | |||| | ||||||||||| || ||
Db        61 ATCTACCACAGACCCCTGGACCGGCCTGCTAGCCCATGCACCGATGTTAATGATATCGAN 120

Qy       422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
               ||| || ||| ||||||| |||||||||||||||||||||||| ||||||||| ||||
Db       121 AGCACTCC-CCCAAGGAAATTTCAACTGCACAACCCCTACTACACCCCAATTCAGCAGGA 179

Qy       482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
             |||||||||||| || ||||||||||||||||| ||| ||||||||||||||||||||||
Db       180 AGCAGTTAGAGCGGTCGTCAGCCAACCTCCCCAGCAGCACTTGGGTTTTCCTGTTGAGAG 239

Qy       542 GGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTA 601
              | ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       240 CGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTA 299

Qy       602 NCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGA 661
              |||||||||||||||| || | ||||||||| ||||||||||||||||||||||||| |
Db       300 GCTGGGAAGGTGACCGCGTCTACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAA 359

Qy       662 CCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAG 712
             ||||||         ||||||||||||||||||||||| |||||||||||||||||| ||
Db       360 CCAATCAGGTAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAGAG 419

Qy       713 CAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAA 772
```

```
               ||||||||||||||||| ||||||||||||||| ||| |||||| |||  |||||||||
Db        420 AAATAGCCAATCATCTATCGCCTGAGAGCACAGCAGGAGGGACAATGATCCGGATATAAA 479

Qy        773 CTCAGGCATTCAAGCCAGCAACAGCAACCCCCCTTTGGGTCCCCTCCCATTGTATGGAGC 832
              | || |||||| |||||||||| ||  |||||||||| |||||||||| |||||||||||
Db        480 CCCAAGCATTCGAGCCAGCAACGGCTGCCCCCCTTTGTGTCCCCTCCCTTTGTATGGAGC 539

Qy        833 TCTGTTTTCACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTT 892
              |||||          ||||||||||||||||| ||||||||||||||||||||| |||||
Db        540 TCTGT----------TTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCTGTGTT 589

Qy        893 TTTTATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACA 952
              | ||| || |  ||||||||||| | ||||| ||||||||||||||||||| || ||  |
Db        590 TGTTACGGTTTGAGCTGAGCTTTCGCTCGCCGTCCACCACTGCTGTTTGCCGCCATCGGA 649

Qy        953 GACCCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCC 1012
              |||| || ||||||||||||||||  |||||  |||| |||| || ||||||||||||||
Db        650 GACCTGCCGCTGACTTCCATCCCTCCGGATCTGGCAGGGTGTTCA-TGTGCTCCTGATCC 708

Qy       1013 AGCGAGGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGC 1072
              || |||| ||||||||||| ||| |||  ||||||||||||||||||||||||||| | |
Db        709 AGAGAGGCACCCATTGCCATTCCTGATTGGGCTAAAGGCTTGCCATTGTTCCTGCANGAC 768

Qy       1073 TAAGTGCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCT 1132
              |||||||| |||||  ||||||| || ||||||||| ||| |||||||||| | ||||||
Db        769 TAAGTGCCCGGGTTCATCCTAATCGAGCTGAACACTAGTCGCTGGGTTCCACGATTCTCT 828

Qy       1133 TCCATGACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATT 1192
              ||| ||||||||| ||||||||||||||||||||||||||||| ||||||||||||||||
Db        829 TCCGTGACCCACGACTTCTAATAGAGCTATAACACTCACCGCACGGCCCAAGATTCCATT 888

Qy       1193 CCTTGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTG 1252
              |||| | ||| ||||||||||||||||||||||||||||  ||||||||||||||||||
Db        889 CCTTCGAATCCGTGAGGCCAAGAACCCCAGGTCAGAGAACATGAGGCTTGCCACCATTTT 948

Qy       1253 GGAAGTGGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCA 1312
              ||||||||||  | |||||||||| || ||||  ||||||||||||||||| ||||||||
Db        949 GGAAGTGGCCTGCCGCCATTTTGGAAGTGGCCTGCCACCATCTTGGGAGCTCTGGGAGCA 1008

Qy       1313 AGGATCCCCCAGTAACA 1329
              |||| ||||  ||||||
Db       1009 AGGACCCCCTGGTAACA 1025

RESULT 10
AAA63826
ID   AAA63826 standard; DNA; 2030 BP.
XX
AC   AAA63826;
XX
DT   06-AUG-2003  (revised)
DT   04-DEC-2000  (first entry)
XX
DE   Nucleotide sequence of the MSRV-1 3' env and LTR regions.
XX
KW   MSRV-1; pol region; long terminal repeat; LTR; RU5 region; retrovirus;
KW   ss.
XX
OS   Multiple sclerosis associated retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1629
FT                   /*tag=  a
FT                   /note= "Contains one termination codon"
FT   sig_peptide     1..81
FT                   /*tag=  b
FT   CAAT_signal     1800..1807
FT                   /*tag=  c
FT   CAAT_signal     1858..1864
FT                   /*tag=  d
FT   TATA_signal     1906..1911
FT                   /*tag=  e
FT   polyA_signal    1996..2002
```

```
FT                   /*tag=  f
XX
PN   WO200047745-A1.
XX
PD   17-AUG-2000.
XX
PF   15-FEB-2000; 2000WO-IB000159.
XX
PR   15-FEB-1999;   99EP-00420041.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Perron H,  Komurian-Pradel F;
XX
DR   WPI; 2000-506097/45.
DR   P-PSDB; AAB08195.
XX
PT   Nucleotide fragment of LTR-RU5 region from Multiple Sclerosis retrovirus
PT   (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological
PT   sample.
XX
PS   Disclosure; Fig 2; 23pp; English.
XX
CC   The present sequence represents the nucleotide sequence corresponding to
CC   the 3' env region and long terminal repeat sequences from clone CL6 of
CC   Multiple Sclerosis retrovirus (MSRV-1). The specification describes a
CC   long terminal repeat (LTR)-RU5 region which encodes the expression of a
CC   MSRV-1 protein. This is unusual for LTRs, in particular in the RU5
CC   region. The sequence includes CAAT and TATA signals which are present in
CC   the U3 and R regions and are not directed towards the CDS indicated in
CC   the features table. Probes and antibodies to the MSRV-1 retrovirus
CC   protein and encoding polynucleotide sequences are used to detect the
CC   presence of MSRV-1 retrovirus in a biological sample. (Updated on 06-AUG-
CC   2003 to correct OS field.)
XX
SQ   Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 U; 0 Other;

  Query Match             58.3%;  Score 775.4;  DB 3;  Length 2030;
  Best Local Similarity   92.8%;  Pred. No. 1.8e-243;
  Matches  812;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 1199

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||| |||||| ||||||| ||||||||||||||  ||||  | |||||||||||||||||
Db       1200 AGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGA 1259

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| |||||||||||||||| ||||||||||| | || |||||||||
Db       1260 AATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCT 1319

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||  |||||||||||||||||||||||||||| |||||||
Db       1320 CCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATT 1379

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               |||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||
Db       1380 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAAT 1439

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              |||||||||||||||||| || ||  | |||||||||||||||||| |||||||||||||
Db       1440 TGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCAGATGGAGTCCATGACTAA 1499

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||| |||||||||||||||||||||||||||| | |||||| |||||||||||||||||
Db       1500 GATCCACCGTGGACCCCTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGA 1559

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||||||||||||||||||||||||||||||||||  | |||||||||  ||
Db       1560 AGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGG 1619
```

```
Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| ||  |||||||||||||||||||| |||||||||||||||||||||
Db       1620 AAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGA 1679

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||| |||||||||||||||||||||||||||||||||||  || ||||||||| ||||||
Db       1680 GGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCT 1739

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | ||||||||||||| ||||||| || |||||||||||||||||||||||||||||||||
Db       1740 AGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCG 1799

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||| ||||||||||| |||||||
Db       1800 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCC 1859

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||| ||||||
Db       1860 AATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCA 1919

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||| ||||||||||
Db       1920 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTT 1979

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              |||||||||||||||||||||||| |||||||| |
Db       1980 CACTCTATTTCACTCTATTAAATCTTGCAACTGAA 2014

RESULT 11
AAD41225
ID   AAD41225 standard; cDNA; 2074 BP.
XX
AC   AAD41225;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Human EMBRY-2 cDNA.
XX
KW   Human; embryogenesis associated protein; AIDS; reproductive disorder;
KW   infertility; endometriosis; endometrial tumour; inflammatory disorder;
KW   autoimmune disorder; acquired immune deficiency syndrome; transgenic;
KW   ovarian tumour; contact dermatitis; placenta disorder; preeclampsia;
KW   EMBRY-2; allergy; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             64..1692
FT                   /*tag=  a
FT                   /product= "EMBRY-2 protein"
FT   sig_peptide     64..123
FT                   /*tag=  b
FT   mat_peptide     124..1689
FT                   /*tag=  c
FT                   /product= "Mature EMBRY-2 protein"
XX
PN   WO200248362-A2.
XX
PD   20-JUN-2002.
XX
PF   14-NOV-2001; 2001WO-US043956.
XX
PR   15-NOV-2000; 2000US-0249407P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Ramkumar J,  Arvizu C;
XX
DR   WPI; 2002-537629/57.
DR   P-PSDB; AAE25054.
XX
PT   New polypeptides of human embryogenesis associated proteins for screening
```

```
PT   modulators useful for treating or preventing disorders e.g.
PT   endometriosis, infertility, allergy, preeclampsia.
XX
PS   Claim 59; Page 96-97; 97pp; English.
XX
CC   The invention relates to human embryogenesis associated proteins (EMBRY)
CC   and nucleic acid molecules encoding such proteins. EMBRY sequences are
CC   useful for screening modulators useful for treating or preventing
CC   disorders associated with abnormal expression of EMBRY. The disorders
CC   treated include reproductive disorders such as infertility,
CC   endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory
CC   disorder such as acquired immune deficiency syndrome (AIDS), allergies,
CC   contact dermatitis; disorders of the placenta such as preeclampsia,
CC   abruptio placentae etc. Sequences of the invention are also useful for
CC   analysing a proteome of a tissue or a cell type. EMBRY proteins are
CC   useful as immunogens for preparing antibodies. Polynucleotides of the
CC   invention are useful for creating knockin humanised animals or transgenic
CC   animals to model human diseases. They are also used in gene therapy. The
CC   present sequence is human EMBRY-2 cDNA
XX
SQ   Sequence 2074 BP; 583 A; 567 C; 392 G; 531 T; 0 U; 1 Other;

  Query Match             57.7%;  Score 766.8;  DB 6;  Length 2074;
  Best Local Similarity   93.0%;  Pred. No. 1.2e-240;
  Matches  812;  Conservative    0;  Mismatches   60;  Indels    1;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 1262

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||| |||||||||||||| ||||||||||||||  ||||  | |||||||||||||||||
Db       1263 AGGAGAAGAATGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGA 1322

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| |||||||||||||||| ||||||||||| | || |||||||||
Db       1323 AATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACACTGGGGCCT 1382

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||  |||||||||||||||||||||||||||| |||||||
Db       1383 CCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATT 1442

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               |||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||
Db       1443 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAAT 1502

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              ||||||||||||||||||||| |||||||||||||| |||||||||||||||||||||||
Db       1503 TGAAGCTGTAAAGCTACAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAA 1562

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||  ||||||||||||||||||||||| | |||||| |||||||||||||| ||
Db       1563 AATCTACCACGGACCCCTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGA 1622

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||| |||||| |||||||||||||||||||||||||||||  | ||||||||| |||
Db       1623 AGGCACTCCTCCCAAGGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCAGG 1682

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||||| ||||||||||||||| |||||||||||||||||||||
Db       1683 AAGCAGTTAGAGCGGTCGTCAGTCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGA 1742

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||| ||||||||||||||||||||||||||||||||||| || |||||||||| ||||||
Db       1743 GGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGATTAAGAATCCCTAAGCCT 1802

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||||| |||| ||||||||| |||||||||||||||||||||||||
Db       1803 AGCTGGGAAGGTGACCGCGTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCA 1862

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       1863 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 1922
```

```
Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||||||||||||||| |||| ||||||||||||  |||||||| ||||||
Db       1923 AATCATCTATTGCCTGAGAGCACAGTGGGAGGGACAAGGATTGCAATATAAACCCAGGCA 1982

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| ||||||| | ||||||| |||||||||||| |||| ||||||||||||||||||||
Db       1983 TTCGAGCCAGC-ANAGCAACCGCCTTTGGGTCCCTTCCCCTTGTATGGGAGCTCTGTTTT 2041

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTG 873
              |||||||||||||||||||||||| ||||||||
Db       2042 CACTCTATTTCACTCTATTAAATCTTGCAACTG 2074


RESULT 12
ACC46747
ID   ACC46747 standard; cDNA; 2046 BP.
XX
AC   ACC46747;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human dithp growth/development-associated protein-encoding cDNA.
XX
KW   Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW   cancer; cell proliferative disorder; autoimmune disorder;
KW   inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW   neurological disorder; gastrointestinal disorder; transport disorder;
KW   connective tissue disorder; drug screening; proteome analysis;
KW   gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW   disease model; toxicological testing; transcript imaging; growth;
KW   development; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200297031-A2.
XX
PD   05-DEC-2002.
XX
PF   27-MAR-2002; 2002WO-US010056.
XX
PR   28-MAR-2001; 2001US-0279619P.
PR   29-MAR-2001; 2001US-0280067P.
PR   29-MAR-2001; 2001US-0280068P.
PR   16-MAY-2001; 2001US-0291280P.
PR   17-MAY-2001; 2001US-0291829P.
PR   17-MAY-2001; 2001US-0291849P.
PR   19-JUN-2001; 2001US-0299428P.
PR   20-JUN-2001; 2001US-0299776P.
PR   20-JUN-2001; 2001US-0300001P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Daffo A,  Jones AL,  Tran AB,  Dahl CR,  Gietzen D,  Chinn J;
PI   Dufour GE,  Hillman JL,  Yu JY,  Tuason O,  Yap PE,  Amshey SR;
PI   Daughtery SC,  Dam TC,  Liu TF,  Nguyen DA,  Kleefeld Y,  Gerstin EH;
PI   Peralta CH,  David MH,  Lewis SA,  Chen AJ,  Panzer SR,  Harris B;
PI   Flores V,  Marwaha R,  Lo A,  Lan RY,  Urashka ME;
XX
DR   WPI; 2003-129518/12.
DR   P-PSDB; ABR41810.
XX
PT   Novel human diagnostic and therapeutic polypeptide useful for identifying
PT   test compound which specifically binds to a polypeptide encoded by human
PT   diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
PS   Claim 2; SEQ ID NO 668; 591pp; English.
XX
CC   The invention relates to novel human diagnostic and therapeutic
CC   polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC   proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC   polynucleotide sequences at least 90% identical to the dithp cDNA
CC   sequences of the invention; recombinant vectors, host cells and
CC   transgenic organisms comprising a dithp nucleic acid sequence; the
CC   recombinant production of DITHP proteins; antibodies specific for DITHP
CC   proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC   detecting dithp nucleotide and protein sequences; methods of screening
CC   for compounds which specifically bind a DITHP protein; and methods of
CC   assessing the toxicity of test compounds using a dithp hybridisation
CC   probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC   diagnosis of a wide variety of conditions including cancer and other cell
CC   proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC   viral, fungal or parasitic infections; hormonal disorders; metabolic
CC   disorders; neurological disorders; gastrointestinal disorders; transport
CC   disorders; and connective tissue disorders. They may also be used to
CC   screen for modulators of protein activity or gene expression. DITHP
CC   proteins can additionally be used in analysis of the proteome of a tissue
CC   or cell type and to induce antibodies. The dithp nucleic acids are
CC   additionally useful in somatic or germline gene therapy of the disorders
CC   mentioned above, as a source of antisense sequences, as a source of
CC   probes and primers, in genotyping and identification of individuals, in
CC   the generation of transgenic animal models of human disease or knock in
CC   humanised animals, in toxicological testing, and in transcript imaging.
CC   The present sequence represents a dithp cDNA encoding a DITHP protein
CC   which is associated with growth and development. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2046 BP; 587 A; 494 C; 391 G; 566 T; 0 U; 8 Other;

  Query Match             56.3%;  Score 747.6;  DB 8;  Length 2046;
  Best Local Similarity   86.5%;  Pred. No. 2.6e-234;
  Matches  885;  Conservative    0;  Mismatches  122;  Indels   16;  Gaps    5;

Qy        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
              | ||||||||| ||| |||||||||||||||||||||||||||||||| ||||||| |||
Db        571 CTCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAACTGTAAAACTA 630

Qy        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
              ||||| |||||||||||||| ||||||||||||||||||||||| ||| |||   |||||
Db        631 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACCACAGACCC 690

Qy        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436
              ||||||| ||||||||| | |||||| |||||||||||||| |||| |||||| ||||||
Db        691 CTGGACCAGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCCCCCGAG 750

Qy        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496
              |||||||||||||  ||||||||||||  | ||||||||| |||||||||||| ||| ||
Db        751 GAAATCTCAACTGTGCAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTACAGCGGT 810

Qy        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556
                |||||||||||||||||||| |||||||||||||||||||||| | | ||||||||||
Db        811 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGAGGGCACTGAGAGAC 870

Qy        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616
              | ||||||||||||||||||||| ||||||||||||| |||| || ||||||||||||| 
Db        871 AAGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCGTAGCTGGGAAGGTGACT 930

Qy        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676
              ||||||| ||||||||||||||||||| ||||||||||||||| ||||||||||||||||
Db        931 GCATCCACCTTTAAACATGGGGCTTGCCACTTAGCTCACACCCAACCAATCAGAGAGCTC 990

Qy        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736
              |||||||||||||| ||||||||||||||||||||| |||||||||||| || |||||||
Db        991 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCACCTTTTGCCTG 1050

Qy        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796
              |||||||||| ||| |||||||||| ||||||||||| | |||||||||||| ||||  |
Db       1051 AGAGCACAGCAGGAGGGACAAGGATCGGGATATAAACCCGGGCATTCAAGCCGGCAATGG 1110

Qy        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856
              ||||||||||||||||||||||| |||||||||||||||||          |||||||||
Db       1111 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT----------TTTCACTCT 1160

Qy        857 ATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTATGGCTCAAGCTGAGCTTTT 916
```

```
              |||||||| |||||||||||||||||||||| ||||| ||| ||||| |||||||||| |
Db       1161 ATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTGTTACGGCTCGAGCTGAGCTTCT 1220

Qy        917 GTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTTCCATCCCT 976
              | | ||  ||||||||||||| | |    | || ||||||||| |||||  |||| | ||
Db       1221 GCTTGCAGTCCACCACTGCTGATGGTTGTCATCGCAGACCCGCGGCTGATGTCCACCACT 1280

Qy        977 TTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCATTGCCACTCCC 1036
                |||||| |||| ||||||||||||||||||||||||| |||  || ||||||| ||||
Db       1281 CCGGATCCGGCAGGGTGTCCACTGTGCTCCTGATCCAGCAAGGCGCCAATTGCCGCTCCT 1340

Qy       1037 GATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTCCTAATA 1096
              |||| ||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1341 GATCGGGCTAAAGGCTTGCCATTGTTCCTGCACGGCTAAGTGCCTGGGTTTGTCCTAATT 1400

Qy       1097 GAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTT-CCATGACCCACGGC-TTCTAAT 1154
              || ||| ||||| ||||||||||||||||||| |||| || ||||||||| | |||||||
Db       1401 GAGCTGGACACTAGTCACTGGGTTCCATGGTTGTCTTGCCGTGACCCACGACGTTCTAAT 1460

Qy       1155 AGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAGGCCAAG 1214
              |||||| |||||||    |   | | ||||||||| |||||| | ||| |||||||||||
Db       1461 AGAGCTGTAACACTGCATG---GCCTCAAGATTCCGTTCCTTAGAATCCGTGAGGCCAAG 1517

Qy       1215 AA-CCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGCCATTT 1273
              || |||||||||||||||   ||||||||||| ||| | |||||||||||   | |  |
Db       1518 AACCCCCAGGTCAGAGAACACGAGGCTTGCCATCATCTTGGAAGTGGCCCCACGACCATC 1577

Qy       1274 TGG 1276
              | |
Db       1578 TTG 1580
```

```
RESULT 13
AAX77526
ID   AAX77526 standard; cDNA; 2946 BP.
XX
AC   AAX77526;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human secreted protein AJ172_2 cDNA.
XX
KW   Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW   bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW   cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW   haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;
KW   haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW   cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
OS   Homo sapiens.
XX
PN   WO9926972-A1.
XX
PD   03-JUN-1999.
XX
PF   17-NOV-1998;   98WO-US024614.
XX
PR   21-NOV-1997;   97US-00976110.
PR   18-MAY-1998;   98US-00080478.
PR   20-OCT-1998;   98US-00175928.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M;
XX
DR   WPI; 1999-357813/30.
DR   P-PSDB; AAY08622.
XX
PT   New polynucleotides encoding secreted proteins.
XX
PS   Claim 13a; Page 100-101; 142pp; English.
XX
CC   This invention describes novel human secreted proteins encoded by
CC   polynucleotides isolated from human adult testes, adult brain, adult
CC   blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC   libraries. The products of the invention are predicted to have biological
CC   activities which would make them suitable for treating, preventing or
CC   ameliorating medical conditions in humans and animals, although no
CC   supporting data is given. Suggested activities include nutritional
CC   activity, cytokine and cell proliferation/differentiation activity,
CC   immune stimulating (e.g. as vaccines) or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, and tumour
CC   inhibition activity. The polynucleotides are also stated to be useful for
CC   gene therapy
XX
SQ   Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

  Query Match             55.1%;  Score 731.8;  DB 2;  Length 2946;
  Best Local Similarity   90.7%;  Pred. No. 5e-229;
  Matches  794;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       2067 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 2126

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       2127 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2186

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       2187 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2246

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2247 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2306

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       2307 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2366

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       2367 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2414

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2415 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2474

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2475 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2534

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2535 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2594

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       2595 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2654

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2655 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2714

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2715 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2774

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2775 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2834
```

```
Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2835 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2894

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2895 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2929


RESULT 14
AAZ59468
ID   AAZ59468 standard; cDNA; 2946 BP.
XX
AC   AAZ59468;
XX
DT   11-APR-2000  (first entry)
XX
DE   Human secreted protein AJ172_2 polynucleotide sequence.
XX
KW   Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW   placental pathology; metastasis inhibition; nutritional activity;
KW   immune stimulator; haematopoiesis regulator; tissue growth;
KW   tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9960020-A1.
XX
PD   25-NOV-1999.
XX
PF   17-MAY-1999;   99WO-US010915.
XX
PR   18-MAY-1998;   98US-00080478.
PR   20-OCT-1998;   98US-00175928.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Mi S,  Treacy M;
XX
DR   WPI; 2000-116311/10.
DR   P-PSDB; AAY67313.
XX
PT   New polynucleotides encoding secreted cDNA libraries, used to develop
PT   products for the diagnosis and treatment of neoplastic disease.
XX
PS   Claim 14; Page 107-108; 149pp; English.
XX
CC   This is the human secreted protein AJ172_2 nucleotide sequence, obtained
CC   from a human adult testes cDNA library. The invention relates to secreted
CC   human and murine proteins. The polynucleotides and proteins are predicted
CC   to have biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals. Detection of the levels of the proteins can be used for the
CC   diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
CC   which modulate the expression or function of the proteins may be used for
CC   treating a neoplastic disease and inhibiting metastasis. Other suggested
CC   activities include nutritional activity (e.g. in feeds), cytokine and
CC   cell proliferation/differentiation activity, immune stimulating (e.g. as
CC   vaccines) or suppressing activity, haematopoiesis regulating activity,
CC   tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC   invasion suppressor activity, and tumour inhibition activity. The
CC   polynucleotide sequences are also stated to be useful for gene therapy
XX
SQ   Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

  Query Match             55.1%;  Score 731.8;  DB 3;  Length 2946;
  Best Local Similarity   90.7%;  Pred. No. 5e-229;
  Matches  794;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;
```

```
Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       2067 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 2126

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       2127 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2186

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       2187 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2246

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2247 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2306

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       2307 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2366

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       2367 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2414

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2415 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2474

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2475 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2534

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2535 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2594

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       2595 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2654

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2655 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2714

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2715 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2774

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2775 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2834

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2835 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2894

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2895 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2929


RESULT 15
ADC38776
ID   ADC38776 standard; cDNA; 2946 BP.
XX
AC   ADC38776;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human cDNA encoding a secreted protein #63.
XX
KW   ss; gene; immune disorder; severe combined immunodeficiency; SCID;
KW   autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW   rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;
KW   lymphoid cell deficiency; osteoporosis; osteoarthritis;
KW   peripheral nervous system disease; peripheral neuropathy;
KW   Alzheimer's disease; Parkinson's disease; coagulation disorder;
KW   inflammatory disease; systemic inflammatory response syndrome; SIRS;
KW   ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;
KW   hypersensitivity; regeneration; neural cell proliferation; fertility;
KW   tumour; chemokine; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   US2002193567-A1.
XX
PD   19-DEC-2002.
XX
PF   02-APR-2002; 2002US-00114893.
XX
PR   11-AUG-1995;   95US-00514014.
PR   05-APR-1996;   96US-00628364.
PR   19-APR-1996;   96US-00635311.
PR   07-JUN-1996;   96US-00659224.
PR   17-JUN-1996;   96US-00664596.
PR   09-JUL-1996;   96US-00677231.
PR   26-JUL-1996;   96US-00686878.
PR   23-AUG-1996;   96US-00701819.
PR   27-SEP-1996;   96US-00721488.
PR   27-SEP-1996;   96US-00721798.
PR   27-SEP-1996;   96US-00721923.
PR   27-SEP-1996;   96US-00721926.
PR   25-OCT-1996;   96US-00738367.
PR   30-OCT-1996;   96US-00739775.
PR   13-JAN-1997;   97US-00783395.
PR   10-APR-1997;   97US-00833823.
PR   02-JUN-1997;   97US-00867677.
PR   05-SEP-1997;   97US-00924838.
PR   06-OCT-1999;   99US-00413232.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR,  Spaulding V,  Carlin-Duckett M;
PI   Kelleher K;
XX
DR   WPI; 2003-657236/62.
DR   P-PSDB; ADC38777.
XX
PT   Proteins AZ3021 encoded by clone AZ3021 from human adult colon, and
PT   BD12716 encoded by clone BD12716 from human fetal kidney cDNA library,
PT   useful for treating e.g. multiple sclerosis and rheumatoid arthritis.
XX
PS   Disclosure; SEQ ID NO 134; 412pp; English.
XX
CC   The invention relates to a protein comprising fully defined AZ302 1
CC   protein or BD127 1 6 protein. The polynucleotides are useful for
CC   expressing recombinant proteins for analysis and are also useful as
CC   chromosome markers or tags to identify chromosomes or to map related gene
CC   positions. The proteins are useful as amino acid supplement, carbon
CC   source, nitrogen source and carbohydrate source. The proteins are useful
CC   for treating various immune deficiencies and disorders (e.g. severe
CC   combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC   sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic
CC   reactions (e.g. asthma), myeloid or lymphoid cell deficiencies,
CC   osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g.
CC   peripheral neuropathy, Alzheimer's disease, Parkinson's disease),
CC   coagulation disorders, inflammatory diseases (e.g. systemic inflammatory
CC   response syndrome (SIRS), ischaemia-reperfusion injury, Crohn's disease),
CC   anaphylaxis and hypersensitivity. Proteins are also useful for inducing
CC   tumour immunity, for inducing bone, cartilage, tendon, ligament and/or
CC   nerve growth or regeneration, for proliferating neural cells and for
CC   regenerating nerve and brain tissue, for inducing fertility and for
CC   inhibiting tumour growth. Proteins are also useful as chemokine for
CC   mammalian cells (e.g., monocytes, fibroblasts, neutrophils), and also
CC   useful as inhibitors of receptor/ligand interactions. The present
CC   sequence represents cDNA encoding a human secreted protein.
XX
SQ   Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

  Query Match             55.1%;  Score 731.8;  DB 10;  Length 2946;
  Best Local Similarity   90.7%;  Pred. No. 5e-229;
  Matches  794;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       2067 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 2126

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       2127 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2186

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       2187 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2246

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2247 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2306

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       2307 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2366

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       2367 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2414

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2415 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2474

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2475 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2534

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2535 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2594

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       2595 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2654

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2655 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2714

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2715 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2774

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2775 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2834

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2835 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2894

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2895 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2929


RESULT 16
AAD24195
ID   AAD24195 standard; cDNA; 2930 BP.
```

```
XX
AC   AAD24195;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Human syncytin cDNA.
XX
KW   Human; syncytin; preeclampsia; gestational trophoblast disorder;
KW   choriocarcinoma; hydatiform mole; placental site tumour; abortion;
KW   envelope gene; human endogenous defective retrovirus; HERV-W; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             930. .2546
FT                   /*tag=  a
FT                   /product= "Syncytin"
XX
PN   WO200204678-A2.
XX
PD   17-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-US021719.
XX
PR   07-JUL-2000; 2000US-0216657P.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Keith JC,  Mccoy JM,  Mi S;
XX
DR   WPI; 2002-171727/22.
DR   P-PSDB; AAE14540.
XX
PT   Identifying a compound for treating a subject with or at risk of
PT   developing preeclampsia, comprises determining whether the expression or
PT   activity of syncytin in the cell is modulated in the presence of a test
PT   compound.
XX
PS   Disclosure; Page 39-42; 43pp; English.
XX
CC   The invention relates to identifying compounds which are modulators of
CC   syncytin expression. The syncytin modulators are useful in diagnosis and
CC   treatment of preeclampsia and gestational trophoblast disorders (e.g.
CC   choriocarcinoma, hydatiform mole, placental site tumour and missed/
CC   incomplete abortion). Syncytin is a human gene derived from the envelope
CC   gene of human endogenous defective retrovirus, HERV-W. The present
CC   invention is based partly on the discovery that syncytin expression is
CC   dramatically reduced in preeclampsia, and is also mis-localised to the
CC   apical syncytiotrophoblast membrane. The present sequence is human
CC   syncytin cDNA
XX
SQ   Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 U; 0 Other;

  Query Match            55.0%;  Score 730.8;  DB 6;  Length 2930;
  Best Local Similarity   90.7%;  Pred. No. 1.1e-228;
  Matches  793;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        2069 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 2128

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        2129 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2188

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        2189 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2248

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        2249 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2308
```

```
Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        2309 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2368

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        2369 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2416

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        2417 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2476

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        2477 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2536

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        2537 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2596

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        2597 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2656

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2657 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2716

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db        2717 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2776

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        2777 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2836

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        2837 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2896

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTGC 874
               ||  |||||||||||||||||||| |||||||||
Db        2897 CATGCTATTTCACTCTATTAAATCTTGCAACTGC 2930


RESULT 17
AAF55630
ID   AAF55630 standard; DNA; 2781 BP.
XX
AC   AAF55630;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
KW   Envelope protein; HERV; syncytia formation; placental development;
KW   syncytia; cancer; cell adhesion; ss.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             762. .2378
FT                   /*tag=  a
FT                   /product= "envelope protein"
XX
PN   WO200116171-A1.
XX
PD   08-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-FR002429.
XX
PR   01-SEP-1999;    99FR-00011141.
PR   15-SEP-1999;    99FR-00011793.
```

```
XX
PA   (INMR ) BIO MERIEUX.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI   Mallet F,  Cosset F,  Blond J,  Lavillette D,  Bouton O,  Ruggieri A;
XX
DR   WPI; 2001-226676/23.
DR   P-PSDB; AAB67652.
XX
PT   Detecting expression of human endogenous retrovirus envelope protein in
PT   cells of a tissue or culture, from its ability to induce syncytia.
XX
PS   Disclosure; Page 44-45; 57pp; French.
XX
CC   The present sequence encodes a human endogenous retrovirus envelope
CC   protein. The specification describes a method for detecting expression of
CC   an envelope protein from a human endogenous retrovirus (HERV), in cells,
CC   of a tissue or culture. The method comprises detecting syncytia formation
CC   due to the fusogenic properties of the envelope protein. Envelope
CC   polypeptides and polynucleotides are used to produce therapeutic or
CC   prophylactic compositions, particularly for treatment of cancer, to
CC   correct defects in placental development (or other natural formation of
CC   other types of syncytia), and to promote adhesion of cells in grafts or
CC   cellular repair processes. Expression of sequences antisense to the
CC   polynucleotide are used to prevent formation of syncytia
XX
SQ   Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 U; 0 Other;

  Query Match             54.9%;  Score 730.2;  DB 5;  Length 2781;
  Best Local Similarity   90.6%;  Pred. No. 1.6e-228;
  Matches  793;  Conservative    0;  Mismatches   70;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1901 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1960

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1961 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2020

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       2021 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2080

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2081 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2140

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       2141 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2200

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       2201 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2248

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2249 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2308

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2309 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2368

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2369 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2428

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       2429 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2488

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2489 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2548

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2549 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2608

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2609 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2668

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| |||||||||||||||||||
Db       2669 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTT 2728

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2729 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2763


RESULT 18
AAS84210
ID   AAS84210 standard; cDNA; 6394 BP.
XX
AC   AAS84210;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #20014.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG20023.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 20014; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6394 BP; 1840 A; 1597 C; 1384 G; 1571 T; 0 U; 2 Other;

  Query Match             54.8%;  Score 728.6;  DB 5;  Length 6394;
  Best Local Similarity   78.6%;  Pred. No. 9.1e-228;
  Matches 1094;  Conservative    0;  Mismatches  223;  Indels   74;  Gaps   16;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||| | ||||||||  |||||||||||||||||||||||||
Db       1175 TCAAAATCGAAGAGCTTTAGACCTTCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1234

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       1235 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1294

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       1295 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1354

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1355 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1414

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       1415 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1474

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       1475 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 1522

Qy        361 AATCTACCGTGGACCCCT-GGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTG 419
               ||||||||  ||||||| ||||||||||||||| | | | |||||||||||||||||
Db       1523 GATCTACCGCAGACCCCTGGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCA 1582

Qy        420 AAGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCC---------- 469
              ||| ||||||||| |||||||||||| ||||| || ||       || ||
Db       1583 AAGGCACCCCTCCTGAGGAAATCTCAGCTGCAAAAACCTCTACTTACGCCCCAATTCAGG 1642

Qy        470 -----AATTCAGTAGGAAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTG 524
                   ||  ||||  | ||  |        |     | |   ||   |   || |    |
Db       1643 CAGGGAAGGCAGTTAGGAGGCGGGTCGTTCGGGCCAACCTTCCCCAACAGGCACTTTAGG 1702

Qy        525 GGTTTTCCTGTTGAGAGGGTGGACTGAGAG----ACAGGACTAGCTGGA--TTTCCTAGG 578
              | |||||||||||     | ||     |||       ||| |||||||   || | ||||
Db       1703 GTTTTTCCTGTTGGAGATGGGGGGACTGAGGAGACAGGGATTAGCTGGGATTTCCTTAGG 1762

Qy        579 CTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACCG------------CATCCATCT 626
              |||||  |  |  ||          | |   || ||  |            |  ||  | |
Db       1763 CTGACCTAAGAATCCCCTAAGGNCTTAGCCTGGGGAAGGGTGGACCACATCCCACCTTTT 1822

Qy        627 TTAAACATGGGG---CTTGCAACTTAGCTCACACCC---GACCAATCAGAGAGCTCACTA 680
              | ||||| ||||    |||||||||      ||| |   |||||||||||||||||||||
Db       1823 TAAAACACGGGGGGCTTTGCAACTTTAGGTTCACACCTTGACCAATCAGAGAGCTCACTA 1882

Qy        681 AAATG-CTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTGA-G 738
              ||||| ||||| ||||||| ||||||||||||| |||||||||||||||||||||||| |
Db       1883 AAATGCCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGG 1942

Qy        739 AGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGC-AACAGC 797
              ||||||| |||| ||||||||||||  |||||||| ||||||||| ||||||| ||| ||
Db       1943 AGCACAGTGGGAGGGACAAGGATTGCAATATAAACCCAGGCATTCGAGCCAGCGAACGGC 2002

Qy        798 AACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGT----------TTTCACTCTA 847
              |||| |||||||||||| |||| |||||||||||||||||          ||||||||||
Db       2003 AACCGCCTTTGGGTCCCTTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCTA 2062

Qy        848 TTTCACTCTATTAAATCATGCAACTG--CACTCTTCTGGTCCGTGTTTTTTATGGCTCAA 905
              ||||||||||||||||| ||||||||  ||||||||||||| ||||||||||  |||| |
Db       2063 TTTCACTCTATTAAATCTTGCAACTGTACACTCTTCTGGTCAGTGTTTTTTACAGCTCTA 2122

Qy        906 GCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCA---CAGACCCGCTGC 962
              ||||||||| || | ||  ||||||||||||||||||| ||| ||   ||||||||||||
Db       2123 GCTGAGCTTCTGCTTGCAGTCCACCACTGCTGTTTGCCGCCGCCACCGCAGACCCGCTGC 2182

Qy        963 TGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTAC 1022
              || |||||||||||| |||||| |||| |||||  |||   ||||||||||||||||  |
Db       2183 TGGCTTCCATCCCTTCGGATCCGGCAGGGTGTCTGCTGCATTCCTGATCCAGCGAGGCGC 2242

Qy       1023 CCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTG 1082
              |||||||||| ||| ||| ||||||||||||||||||||||||||| || |||||||| |
Db       2243 CCATTGCCACGCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCACGGTTAAGTGCCCG 2302

Qy       1083 GGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCC 1142
              |||| ||| |||| || ||||||||| |||||||||||||| ||||||||||| ||||||
Db       2303 GGTTGGTCGTAATCGAGCTGAACACTTGTCACTGGGTTCCACGGTTCTCTTCCGTGACCC 2362

Qy       1143 ACGGCTTCTAATAG--AGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              | ||||||||||||  |    ||||| |  |     |||||||||||||||||||||| |
Db       2363 ATGGCTTCTAATAGGAACTATTAACATTACCACATGGGCCCAAGATTCCATTCCTTGGGA 2422

Qy       1201 --TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGT 1258
                || ||||||||||||||||||||||||||||     ||||||||||||| | |||||
Db       2423 ATTCCGTGAGGCCAAGAACCCCAGGTCAGAGAATACAAGGCTTGCCACCATCTTGGAAGC 2482

Qy       1259 GGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATC 1318
              |||||   ||||| |||| ||||||   ||||||||||||||||| || |  |||||| |
Db       2483 GGCCCGTCGCCATCTTGGAAGCGGCTTGCCACCATCTTGGGAGCTCTGTGCACAAGGACC 2542

Qy       1319 CCCCAGTAACA 1329
              |||| ||||||
Db       2543 CCCCGGTAACA 2553


RESULT 19
AAX25665
ID   AAX25665 standard; cDNA to mRNA; 7582 BP.
XX
AC   AAX25665;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Complete human endogenous retrovirus W genome.
XX
KW   Clone; human endogenous retrovirus; genome; autoimmune disease;
KW   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW   disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9902696-A1.
XX
PD   21-JAN-1999.
XX
PF   06-JUL-1998;   98WO-FR001442.
XX
PR   07-JUL-1997;   97FR-00008815.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Beseme F,  Blond J,  Bouton O,  Mandrand B,  Mallet F;
XX
DR   WPI; 1999-120897/10.
XX
PT   New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT   exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT   disease, and abnormal or failed pregnancy.
XX
PS   Claim 1; Page 71-74; 106pp; French.
XX
```

```
CC   This sequence represents the complete sequence of the human endogenous
CC   retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC   peptides encoded by them are markers of autoimmune disease (e.g. multiple
CC   sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC   insulin- dependent diabetes and related pathologies) and of abnormal or
CC   unsuccessful pregnancy and can be used as chromosomal markers for
CC   susceptibility to these conditions, or proximity markers of genes
CC   associated with this susceptibility
XX
SQ   Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 Other;

  Query Match             54.4%;  Score 722.6;  DB 2;  Length 7582;
  Best Local Similarity   88.5%;  Pred. No. 9.4e-226;
  Matches  774;  Conservative   19;  Mismatches   70;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||:||||||||||||||||||||:|||||||||  |:|||||||||||||||||||||||
Db       6720 TCRAAATCGAAGAGCTTTAGACTYGCTAACCGCTGARAGAGGGGGAACCTGTTTATTTTT 6779

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | ||||||||||||:||||
Db       6780 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTRAAGA 6839

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| :|||||||||| || |||:||||||:|| |||| |||||||| ||||||||||||
Db       6840 AATTCSAGATCGAATACAACGTAKAGCAGARGAGCTTCGAAACACTGGACCCTGGGGCCT 6899

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||:|||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       6900 CCTCAGCCRATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 6959

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| :|| |||||||||| |||||||||||||||||
Db       6960 GCTACTCCTCTTTGGACCCTGTATCTTTRACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 7019

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||:|| |||            |||||||| ||| |||||||||||| ||||||
Db       7020 CGAAGCTGTRAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 7067

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  ||||||||||||||||||:||| | | | |||||||||||||||||  |
Db       7068 GATCTACCGCAGACCCCTGGACCGGCCTGYTAGCCCACGATCTGATGTTAATGACATCAA 7127

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       7128 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 7187

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       7188 AAGCAGTTAGAGCGGTSGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 7247

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||:|||||||||: ||||||
Db       7248 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGAYTAAGAATCCYTAAGCCT 7307

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | :||||||||||||| |||||| ||||||||| ||||||||||||||||:||||||| |
Db       7308 AGSTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGYTCACACCTG 7367

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       7368 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 7427

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||:||||||:|||||||||||| ||| |||||| ||| ||||||||||| || |
Db       7428 AATCATYTATTGCMTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTY 7487

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       7488 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTT 7547

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||:||||:
Db       7548 CATGCTATTTCACTCTATTAAATCTTGCARCTGCR 7582


RESULT 20
AAA59215
ID   AAA59215 standard; DNA; 7582 BP.
XX
AC   AAA59215;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Human endogenous retrovirus W (HERV-W) sequence.
XX
KW   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW   gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   LTR             1. .120
FT                   /*tag=  a
FT                   /note= "R of 5' LTR"
FT   LTR             121. .575
FT                   /*tag=  b
FT                   /note= "U5 of 5' LTR"
FT   primer_bind     579. .596
FT                   /*tag=  c
FT   CDS             5581. .7194
FT                   /*tag=  d
FT                   /note= "ORF1 env538"
FT   CDS             7039. .7194
FT                   /*tag=  e
FT                   /note= "ORF2 52 AA"
FT   CDS             7112. .7255
FT                   /*tag=  f
FT                   /note= "ORF3 48 AA"
FT   misc_feature    7244. .7254
FT                   /*tag=  g
FT                   /note= "polypurine tract"
FT   LTR             7256. .7582
FT                   /*tag=  h
FT                   /note= "U3-R of 3' LTR"
FT   polyA_signal    7563. .7569
FT                   /*tag=  i
XX
PN   WO200043521-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-FR000144.
XX
PR   21-JAN-1999;   99FR-00000888.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Mallet F,  Voisset C;
XX
DR   WPI; 2000-499229/44.
XX
PT   New nucleic acid from human endogenous retrovirus, useful e.g. for
PT   diagnosis of autoimmune disease and complications of pregnancy, contains
PT   at least part of the gag gene.
XX
PS   Disclosure; Page 49-52; 53pp; French.
XX
CC   The present sequence represents an endogenetic retrovirus, which is
CC   associated with an autoimmune disease, and is integrated into the human
CC   genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
CC   HERV-W retrovirus is associated with autoimmune disease, failure of
CC   pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
CC   proteins derived from it, are useful for diagnosis of autoimmune disease
CC   (specifically multiple sclerosis) and for monitoring pregnancy. The
CC   nucleic acid fragments may also be used for in situ labelling of isolated
CC   chromosomes, while the transcription product can be used to study or
CC   monitor T cell proliferation in vitro
XX
SQ   Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 0 U; 216 Other;

  Query Match             54.4%;  Score 722.6;  DB 3;  Length 7582;
  Best Local Similarity   88.5%;  Pred. No. 9.4e-226;
  Matches  774;  Conservative   19;  Mismatches   70;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               ||:||||||||||||||||||||:|||||||||  |:|||||||||||||||||||||||
Db        6720 TCRAAATCGAAGAGCTTTAGACTYGCTAACCGCTGARAGAGGGGGAACCTGTTTATTTTT 6779

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | ||||||||||||:||||
Db        6780 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTRAAGA 6839

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| :|||||||||| || |||:||||||:|| |||| |||||||| ||||||||||||
Db        6840 AATTCSAGATCGAATACAACGTAKAGCAGARGAGCTTCGAAACACTGGACCCTGGGGCCT 6899

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||:|||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        6900 CCTCAGCCRATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 6959

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| :|| |||||||||| |||||||||||||||||
Db        6960 GCTACTCCTCTTTGGACCCTGTATCTTTRACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 7019

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||:|| |||            |||||||| ||| |||||||||||| ||||||
Db        7020 CGAAGCTGTRAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 7067

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  ||||||||||||||||||:||| | | | |||||||||||||||||  |
Db        7068 GATCTACCGCAGACCCCTGGACCGGCCTGYTAGCCCACGATCTGATGTTAATGACATCAA 7127

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        7128 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 7187

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        7188 AAGCAGTTAGAGCGGTSGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 7247

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||:|||||||||: ||||||
Db        7248 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGAYTAAGAATCCYTAAGCCT 7307

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | :||||||||||||| |||||| ||||||||| ||||||||||||||||:||||||| |
Db        7308 AGSTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGYTCACACCTG 7367

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        7368 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 7427

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               ||||||:||||||:|||||||||||| ||| |||||| ||| ||||||||||| || |
Db        7428 AATCATYTATTGCMTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTY 7487

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        7488 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 7547

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
               ||  |||||||||||||||||||| ||||:||||:
Db        7548 CATGCTATTTCACTCTATTAAATCTTGCARCTGCR 7582


RESULT 21
AAX25660
ID   AAX25660 standard; cDNA to mRNA; 1136 BP.
XX
AC   AAX25660;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Human endogenous retrovirus W clone cl.C4C5.
XX
KW   Clone; human endogenous retrovirus; genome; autoimmune disease;
KW   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW   disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9902696-A1.
XX
PD   21-JAN-1999.
XX
PF   06-JUL-1998;   98WO-FR001442.
XX
PR   07-JUL-1997;   97FR-00008815.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Beseme F,  Blond J,  Bouton O,  Mandrand B,  Mallet F;
XX
DR   WPI; 1999-120897/10.
XX
PT   New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT   exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT   disease, and abnormal or failed pregnancy.
XX
PS   Claim 1; Page 59-60; 106pp; French.
XX
CC   This sequence represents clone cl.C4C5 of the human endogenous retrovirus
CC   (HERV) W genome. The nucleic acids, their fragments or peptides encoded
CC   by them are markers of autoimmune disease (e.g. multiple sclerosis,
CC   rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC   dependent diabetes and related pathologies) and of abnormal or
CC   unsuccessful pregnancy and can be used as chromosomal markers for
CC   susceptibility to these conditions, or proximity markers of genes
CC   associated with this susceptibility
XX
SQ   Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 U; 0 Other;

  Query Match             54.3%;  Score 721.2;  DB 2;  Length 1136;
  Best Local Similarity   90.0%;  Pred. No. 8.7e-226;
  Matches  787;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               ||||||||||||||||||||||| |||||||||  |||||||||||||||||||||||||
Db         254 TCAAAATCGAAGAGCTTTAGACTCGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 313

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db         314 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 373

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || |||||||||| || |||| |||||||| ||||||||||||
Db         374 AATTCGAGATCGAATACAACGTAGAGCAGAAGAGCTTCGAAACACTGGACCCTGGGGCCT 433

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         434 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 493

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db         494 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 553

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db         554 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 601

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
```

```
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        602 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 661

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        662 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 721

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        722 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 781

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        782 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 841

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||| ||||||| |
Db        842 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGTTCACACCTG 901

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        902 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 961

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||| |||||||||||| ||| |||||| ||| ||||||||||| || |
Db        962 AATCATCTATTGCATGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 1021

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       1022 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 1081

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGC 874
              ||  |||||||||||||||||||| |||| ||||
Db       1082 CATGCTATTTCACTCTATTAAATCTTGCAGCTGC 1115

RESULT 22
AAA59210
ID   AAA59210 standard; DNA; 1136 BP.
XX
AC   AAA59210;
XX
DT   07-NOV-2000  (first entry)
XX
DE   3' pol gene and 3' non coding sequences of HERV-W from human genome.
XX
KW   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW   gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200043521-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-FR000144.
XX
PR   21-JAN-1999;   99FR-00000888.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Mallet F,  Voisset C;
XX
DR   WPI; 2000-499229/44.
XX
PT   New nucleic acid from human endogenous retrovirus, useful e.g. for
PT   diagnosis of autoimmune disease and complications of pregnancy, contains
PT   at least part of the gag gene.
XX
PS   Disclosure; Page 46; 53pp; French.
XX
CC   The present sequence represents an endogenetic retroviral nucleic acid
CC   fragment, which is associated with an autoimmune disease, and is
CC   integrated into the human genome. The fragment is originally derived from
CC   a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W
CC   retrovirus is associated with autoimmune disease, failure of pregnancy or
CC   disorders of pregnancy. The nucleic acid fragment, or proteins derived
CC   from it, are useful for diagnosis of autoimmune disease (specifically
CC   multiple sclerosis) and for monitoring pregnancy. The nucleic acid
CC   fragments may also be used for in situ labelling of isolated chromosomes,
CC   while the transcription product can be used to study or monitor T cell
CC   proliferation in vitro
XX
SQ   Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 U; 0 Other;

  Query Match             54.3%;  Score 721.2;  DB 3;  Length 1136;
  Best Local Similarity   90.0%;  Pred. No. 8.7e-226;
  Matches  787;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||| |||||||||  |||||||||||||||||||||||||
Db        254 TCAAAATCGAAGAGCTTTAGACTCGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 313

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        314 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 373

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || |||||||||| || |||| |||||||| ||||||||||||
Db        374 AATTCGAGATCGAATACAACGTAGAGCAGAAGAGCTTCGAAACACTGGACCCTGGGGCCT 433

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        434 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 493

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        494 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 553

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        554 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 601

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        602 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 661

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        662 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 721

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        722 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 781

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        782 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 841

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||| ||||||| |
Db        842 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGTTCACACCTG 901

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        902 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 961

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||| |||||||||||| ||| |||||| ||| ||||||||||| || |
Db        962 AATCATCTATTGCATGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 1021

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       1022 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 1081

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGC 874
```

```
              ||  |||||||||||||||||||| |||| ||||
Db       1082 CATGCTATTTCACTCTATTAAATCTTGCAGCTGC 1115


RESULT 23
AAX25661
ID   AAX25661 standard; cDNA to mRNA; 2782 BP.
XX
AC   AAX25661;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Human endogenous retrovirus W clone cl.PH74.
XX
KW   Clone; human endogenous retrovirus; genome; autoimmune disease;
KW   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW   disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS   Human endogenous retrovirus.
XX
PN   WO9902696-A1.
XX
PD   21-JAN-1999.
XX
PF   06-JUL-1998;   98WO-FR001442.
XX
PR   07-JUL-1997;   97FR-00008815.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Beseme F,  Blond J,  Bouton O,  Mandrand B,  Mallet F;
XX
DR   WPI; 1999-120897/10.
XX
PT   New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT   exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT   disease, and abnormal or failed pregnancy.
XX
PS   Claim 1; Page 60-63; 106pp; French.
XX
CC   This sequence represents clone cl.PH74 of the human endogenous retrovirus
CC   (HERV) W genome. The nucleic acids, their fragments or peptides encoded
CC   by them are markers of autoimmune disease (e.g. multiple sclerosis,
CC   rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC   dependent diabetes and related pathologies) and of abnormal or
CC   unsuccessful pregnancy and can be used as chromosomal markers for
CC   susceptibility to these conditions, or proximity markers of genes
CC   associated with this susceptibility
XX
SQ   Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;

  Query Match             53.6%;  Score 712.6;  DB 2;  Length 2782;
  Best Local Similarity   89.4%;  Pred. No. 1e-222;
  Matches  782;  Conservative    0;  Mismatches   81;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              || ||||||||||||||||||||||||||||||  | |||||||||||||||||||||||
Db       1902 TCGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTT 1961

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||| ||||
Db       1962 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGA 2021

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  |||||||||| || ||| ||||||||| |||| |||||||| ||||||||||||
Db       2022 AATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2081

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              |||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2082 CCTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2141

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||||||||||||||||||||||||  || |||||||||| |||||||||||||||||
Db       2142 GCTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2201

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               |||||||| || |||            |||||||| ||| |||||||||||| ||||||
Db       2202 CGAAGCTGTGAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2249

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2250 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2309

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2310 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2369

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2370 AAGCAGTTAGAGCGGTGGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2429

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db       2430 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCT 2489

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              |  ||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2490 AGGTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2549

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2550 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2609

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||| ||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2610 AATCATTTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTT 2669

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2670 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2729

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2730 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2764


RESULT 24
AAA59211
ID   AAA59211 standard; DNA; 2782 BP.
XX
AC   AAA59211;
XX
DT   07-NOV-2000  (first entry)
XX
DE   5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.
XX
KW   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW   gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200043521-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-FR000144.
XX
PR   21-JAN-1999;   99FR-00000888.
XX
PA   (INMR ) BIO MERIEUX.
XX
PI   Paranhos-Baccala G,  Mallet F,  Voisset C;
XX
DR   WPI; 2000-499229/44.
XX
PT   New nucleic acid from human endogenous retrovirus, useful e.g. for
PT   diagnosis of autoimmune disease and complications of pregnancy, contains
PT   at least part of the gag.gene.
XX
PS   Disclosure; Page 46-47; 53pp; French.
XX
CC   The present sequence represents an endogenetic retroviral nucleic acid
CC   fragment, which is associated with an autoimmune disease, and is
CC   integrated into the human genome. The fragment is originally derived from
CC   a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W
CC   retrovirus is associated with autoimmune disease, failure of pregnancy or
CC   disorders of pregnancy. The nucleic acid fragment, or proteins derived
CC   from it, are useful for diagnosis of autoimmune disease (specifically
CC   multiple sclerosis) and for monitoring pregnancy. The nucleic acid
CC   fragments may also be used for in situ labelling of isolated chromosomes,
CC   while the transcription product can be used to study or monitor T cell
CC   proliferation in vitro
XX
SQ   Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;

  Query Match            53.6%;  Score 712.6;  DB 3;  Length 2782;
  Best Local Similarity  89.4%;  Pred. No. 1e-222;
  Matches  782;  Conservative    0;  Mismatches   81;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              || ||||||||||||||||||||||||||||||   | |||||||||||||||||||||||
Db       1902 TCGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTT 1961

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||| ||||
Db       1962 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGA 2021

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  |||||||||| || ||| ||||||||| |||| |||||||| ||||||||||||
Db       2022 AATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2081

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              |||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2082 CCTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2141

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||||||||||||||||||||||||  || |||||||||| |||||||||||||||||
Db       2142 GCTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2201

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               |||||||| || |||            |||||||| ||| |||||||||||| ||||||
Db       2202 CGAAGCTGTGAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2249

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  ||||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2250 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2309

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || |||||||||| ||||||||||| |||||||||| |||||||| | |||||||| |||
Db       2310 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2369

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||||| |||| ||||||||||||||||
Db       2370 AAGCAGTTAGAGCGGTGGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2429

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db       2430 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCT 2489

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              |  ||||||||||||| ||||||| |||||||||| |||||||||||||||||||||| |
Db       2490 AGGTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2549

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| |||||||||||||| |||||||
Db       2550 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2609

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||| ||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
```

```
Db       2610 AATCATTTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTT 2669

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| ||||| |||| ||||||||||||||||||||||||| ||||||||||||||||||||
Db       2670 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2729

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2730 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2764


RESULT 25
AAH20069
ID   AAH20069 standard; DNA; 2782 BP.
XX
AC   AAH20069;
XX
DT   08-AUG-2001  (first entry)
XX
DE   HERV-W envelope protein G encoding nucleic acid.
XX
KW   Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW   envelope protein; multiple sclerosis-related superantigen; vaccine;
KW   surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW   antisense-therapy; autoimmune disorder; ds.
XX
OS   Human endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1. .762
FT                   /*tag=  a
FT   CDS             763. .2379
FT                   /*tag=  b
FT                   /product= "HERV-W envelope protein G"
FT                   /transl_except= (pos:790. .792,aa:Phe)
FT                   /transl_except= (pos:793. .795,aa:Thr)
FT                   /transl_except= (pos:812. .814,aa:Leu)
FT                   /transl_except= (pos:818. .820,aa:Ser)
FT                   /transl_except= (pos:862. .864,aa:Tyr)
FT                   /transl_except= (pos:865. .867,aa:Gln)
FT                   /transl_except= (pos:1174. .1176,aa:Arg)
FT                   /transl_except= (pos:1441. .1443,aa:Leu)
FT                   /transl_except= (pos:1903. .1905,aa:Gln)
FT                   /transl_except= (pos:2017. .2019,aa:Lys)
FT                   /transl_except= (pos:2026. .2028,aa:Arg)
FT                   /transl_except= (pos:2044. .2046,aa:Arg)
FT                   /transl_except= (pos:2089. .2091,aa:Gln)
FT                   /transl_except= (pos:2170. .2172,aa:Asn)
FT   3'UTR           2380. .2782
FT                   /*tag=  c
XX
PN   WO200131021-A1.
XX
PD   03-MAY-2001.
XX
PF   30-OCT-2000; 2000WO-EP010659.
XX
PR   28-OCT-1999;   99EP-00402690.
XX
PA   (UYGE-) UNIV GENEVE.
XX
PI   Conrad B,  Mach B;
XX
DR   WPI; 2001-316336/33.
DR   P-PSDB; AAB75138.
XX
PT   New human retrovirus HERV-W ENV proteins/peptides having superantigen
PT   activity useful for diagnosing and treating multiple sclerosis.
XX
PS   Claim 13; Fig 9; 94pp; English.
XX
CC   On the basis of the PBS t-RNA motif used for the classification of human
CC   endogenous retrovirus (HERVs) the full length endogenous provirus which
CC   was been located on the long arm of human chromosome 7 (7q21-22) has been
CC   designated HERV-W. The present invention describes proteins or peptides
CC   (I) having superantigen (SAg) activity comprising the ENV protein (ENV)
CC   of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC   have neuroprotective activity, and can be used in: vaccines; antisense-
CC   therapy; and HERV-W SAg activity-inhibitors. (I) and encoding DNA/RNA are
CC   useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC   disorders. (I) are also useful for identifying substances (and optionally
CC   recovering) capable of binding to a retroviral superantigen associated
CC   with MS, substances capable of blocking SAg activity and substances
CC   capable of blocking transcription or translation of HERV-W retroviral
CC   superantigen. A protein or peptide derived from (I), modified to be
CC   devoid of SAg activity and being capable of generating an immune response
CC   against HERV-W retroviral SAg is useful in therapy. Nucleic acid
CC   molecules encoding (I) are useful as vaccines against MS. Substances
CC   capable of blocking SAg activity, capable of binding to a retroviral
CC   superantigen associated with MS, or capable of blocking transcription or
CC   translation of HERV-W retroviral superantigen for use in treating or
CC   preventing MS, obtained using (I) are useful for the treatment and
CC   prevention of MS. (I) and nucleic acids encoding them are useful for
CC   diagnosing autoimmune disease. The present sequence encodes the
CC   specifically claimed envelope protein of HERV-W designated G
XX
SQ   Sequence 2782 BP; 741 A; 768 C; 564 G; 709 T; 0 U; 0 Other;

  Query Match             53.6%;  Score 712.6;  DB 5;  Length 2782;
  Best Local Similarity   89.4%;  Pred. No. 1e-222;
  Matches  782;  Conservative    0;  Mismatches   81;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              || ||||||||||||||||||||||||||||||  | |||||||||||||||||||||||
Db       1902 TCGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTT 1961

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||| ||||
Db       1962 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGA 2021

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  |||||||||| || ||| ||||||||| |||| |||||||| ||||||||||||
Db       2022 AATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2081

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              |||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2082 CCTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2141

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||||||||||||||||||||||||  || |||||||||| |||||||||||||||||
Db       2142 GCTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2201

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               |||||||| || |||            |||||||| ||| |||||||||||| ||||||
Db       2202 CGAAGCTGTGAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2249

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2250 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2309

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2310 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2369

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2370 AAGCAGTTAGAGCGGTGGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2429

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db       2430 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCT 2489

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              |  ||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2490 AGGTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2549
```

```
Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2550 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2609

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||| ||||||||||||||||||| ||| |||||| ||| ||||||||||| || |  
Db       2610 AATCATTTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTT 2669

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2670 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2729

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2730 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2764


RESULT 26
AAI14608
ID   AAI14608 standard; DNA; 1894 BP.
XX
AC   AAI14608;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #4541 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 4541; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

  Query Match             53.0%;  Score 704.8;  DB 4;  Length 1894;
  Best Local Similarity   89.7%;  Pred. No. 3e-220;
  Matches  770;  Conservative    0;  Mismatches   79;  Indels    9;  Gaps    1;
```

```
Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||| |||||||||| |||||||||
Db        961 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGGGAACCTTTTTATTTTT 1020

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              || |||| |||||||||  |||||||||||||| |||||||| || ||||||||||||||
Db       1021 AGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCACCGAGAAAGTTAAAGA 1080

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  || |||||||||| |||||||| |||| || |||||||||| ||||||||||||
Db       1081 AATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACACTGGACCCTGGGGCCT 1140

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1141 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1200

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               |||||||||||||||||||||||||| ||| |||||||||||||||||| |||||||||
Db       1201 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAAT 1260

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||| ||||| |||||||| |||||||||||||| ||||||||||||||||||| |||
Db       1261 CGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAGTAA 1320

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||  ||||||||||||||||||||||| | |||||||||||||||||||||  |
Db       1321 AATCTACCACGGACCCCTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAA 1380

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||||||||||||||||||||||||||| ||||||| | ||||||||| |||
Db       1381 AGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1440

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||  |||||  ||||||||||||||||| | |||||||||||||||||||
Db       1441 AAGCAGTTAGAGTGGTTGTTGGCCAACCTCCCCAACAGCAGTTGGGTTTTCCTGTTGAGA 1500

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||| |||||||||||||||| || || |||||||||| |  |||||||||||| ||| ||
Db       1501 GGGGGGACTGAGAGACAGGAATAACTAGATTTCCTAGACCAACTAAGAATCCCTAAGACT 1560

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||||| |||| |||||||||  |||||||||||||||||||| |||
Db       1561 AGCTGGGAAGGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCA 1620

Qy        661 ACCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAA 711
              |||||||         ||||||||||||||||||||||| ||||||||||||||| ||||
Db       1621 ACCAATCAGATACTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGATAAA 1680

Qy        712 GCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAA 771
              | ||||||||||||||| |||||||| |||||||| ||| |||||| ||| |||||||||
Db       1681 GAAATAGCCAATCATCTGTTGCCTGACAGCACAGCAGGAGGGACAATGATCGGGATATAA 1740

Qy        772 ACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAG 831
              || ||||||||| ||||||| ||||| |||| |||||||||||||||| |||||||||||
Db       1741 ACCCAGGCATTCGAGCCAGCTACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAG 1800

Qy        832 CTCTGTTTTCACTCTATT 849
              |||||| |||||||||||
Db       1801 CTCTGTCTTCACTCTATT 1818

RESULT 27
ABA56337
ID   ABA56337 standard; DNA; 1894 BP.
XX
AC   ABA56337;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #4642.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 1; SEQ ID NO 4642; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

  Query Match             53.0%;  Score 704.8;  DB 4;  Length 1894;
  Best Local Similarity   89.7%;  Pred. No. 3e-220;
  Matches  770;  Conservative    0;  Mismatches   79;  Indels    9;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||| |||||||||| |||||||||
Db        961 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGGGAACCTTTTTATTTTT 1020

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              || |||| |||||||||  |||||||||||||| |||||||| || ||||||||||||||
Db       1021 AGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCACCGAGAAAGTTAAAGA 1080

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  || |||||||||| |||||||| |||| || |||||||||| ||||||||||||
Db       1081 AATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACACTGGACCCTGGGGCCT 1140

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1141 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1200

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               |||||||||||||||||||||||||| ||| |||||||||||||||||| |||||||||
Db       1201 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAAT 1260

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||| ||||| |||||||| |||||||||||||| ||||||||||||||||||| |||
Db       1261 CGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAGTAA 1320

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||  ||||||||||||||||||||||| | |||||||||||||||||||||  |
Db       1321 AATCTACCACGGACCCCTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAA 1380

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||||||||||||||||||||||||||| ||||||| | ||||||||| |||
Db       1381 AGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1440

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||  |||||  ||||||||||||||||| | |||||||||||||||||||
Db       1441 AAGCAGTTAGAGTGGTTGTTGGCCAACCTCCCCAACAGCAGTTGGGTTTTCCTGTTGAGA 1500

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||| |||||||||||||||| || || |||||||||| |  |||||||||||| ||| ||
Db       1501 GGGGGGACTGAGAGACAGGAATAACTAGATTTCCTAGACCAACTAAGAATCCCTAAGACT 1560

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||||| |||| |||||||||  |||||||||||||||||||| |||
Db       1561 AGCTGGGAAGGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCA 1620

Qy        661 ACCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAA 711
              |||||||         ||||||||||||||||||||||| ||||||||||||||| ||||
Db       1621 ACCAATCAGATACTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGATAAA 1680

Qy        712 GCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAA 771
              | ||||||||||||||| |||||||| |||||||| ||| |||||| ||| |||||||||
Db       1681 GAAATAGCCAATCATCTGTTGCCTGACAGCACAGCAGGAGGGACAATGATCGGGATATAA 1740

Qy        772 ACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAG 831
              || ||||||||| ||||||| ||||| |||| |||||||||||||||| |||||||||||
Db       1741 ACCCAGGCATTCGAGCCAGCTACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAG 1800

Qy        832 CTCTGTTTTCACTCTATT 849
              |||||| |||||||||||
Db       1801 CTCTGTCTTCACTCTATT 1818


RESULT 28
AAI35980
ID   AAI35980 standard; DNA; 1894 BP.
XX
AC   AAI35980;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #4666 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000663.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human placenta.
XX
PS   Claim 25; SEQ ID NO 4666; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
```

```
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

  Query Match             53.0%;  Score 704.8;  DB 4;  Length 1894;
  Best Local Similarity   89.7%;  Pred. No. 3e-220;
  Matches  770;  Conservative    0;  Mismatches   79;  Indels    9;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||| |||||||||| |||||||||
Db        961 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGGGAACCTTTTTATTTTT 1020

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              || |||| |||||||||  |||||||||||||| |||||||| || ||||||||||||||
Db       1021 AGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCACCGAGAAAGTTAAAGA 1080

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  || |||||||||| |||||||| |||| || |||||||||| ||||||||||||
Db       1081 AATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACACTGGACCCTGGGGCCT 1140

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1141 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1200

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               |||||||||||||||||||||||||| ||| |||||||||||||||||| |||||||||
Db       1201 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAAT 1260

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||| ||||| |||||||| |||||||||||||| ||||||||||||||||||| |||
Db       1261 CGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAGTAA 1320

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||  ||||||||||||||||||||||| | |||||||||||||||||||||  |
Db       1321 AATCTACCACGGACCCCTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAA 1380

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||||||||||||||||||||||||||| ||||||| | ||||||||| |||
Db       1381 AGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1440

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||  |||||  ||||||||||||||||| | |||||||||||||||||||
Db       1441 AAGCAGTTAGAGTGGTTGTTGGCCAACCTCCCCAACAGCAGTTGGGTTTTCCTGTTGAGA 1500

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||| |||||||||||||||| || || |||||||||| |  |||||||||||| ||| ||
Db       1501 GGGGGGACTGAGAGACAGGAATAACTAGATTTCCTAGACCAACTAAGAATCCCTAAGACT 1560

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||||| |||| |||||||||  |||||||||||||||||||| |||
Db       1561 AGCTGGGAAGGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCA 1620

Qy        661 ACCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAA 711
              |||||||         ||||||||||||||||||||||| ||||||||||||||| ||||
Db       1621 ACCAATCAGATACTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGATAAA 1680

Qy        712 GCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAA 771
              | ||||||||||||||| |||||||| |||||||| ||| |||||| ||| |||||||||
Db       1681 GAAATAGCCAATCATCTGTTGCCTGACAGCACAGCAGGAGGGACAATGATCGGGATATAA 1740

Qy        772 ACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAG 831
              || ||||||||| ||||||| ||||| |||| |||||||||||||||| |||||||||||
Db       1741 ACCCAGGCATTCGAGCCAGCTACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAG 1800

Qy        832 CTCTGTTTTCACTCTATT 849
              |||||| |||||||||||
Db       1801 CTCTGTCTTCACTCTATT 1818


RESULT 29
ABA45822
```

```
ID   ABA45822 standard; DNA; 1894 BP.
XX
AC   ABA45822;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human breast cell single exon nucleic acid probe #4517.
XX
KW   Human; microarray; single exon probe; gene expression; breast; disease;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157271-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000662.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-496933/54.
XX
PT   New spatially-addressable set of single exon nucleic acid probes, useful
PT   for measuring gene expression in sample derived from human breast,
PT   comprises number of single exon nucleic acid probes.
XX
PS   Claim 1; SEQ ID NO 4517; 327pp + Sequence Listing; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human breast and BT 474 cells. The method involves contacting the
CC   probes with a collection of detectably labelled nucleic acids derived
CC   from mRNA of human breast, and then measuring the label bound to each
CC   probe of the microarray. The probes are useful for verifying the
CC   expression of regions of genomic DNA predicted to encode proteins. They
CC   are useful for gene discovery, and for determining predisposition and/or
CC   prognosing breast disease. Gene expression analysis is useful for
CC   assessing the toxicity of chemical agents on cells. The microarray of
CC   this invention presents a far greater diversity of probes for measuring
CC   gene expression, with far less bias than expressed sequence tag
CC   microarrays. The method is suitable for rapid production of functional
CC   information from genomic sequence. The present sequence is a single exon
CC   nucleic acid probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

  Query Match             53.0%;  Score 704.8;  DB 4;  Length 1894;
  Best Local Similarity   89.7%;  Pred. No. 3e-220;
  Matches  770;  Conservative    0;  Mismatches   79;  Indels    9;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||| |||||||||| |||||||||
Db         961 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGGGAACCTTTTTATTTTT 1020

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               || |||| |||||||||  |||||||||||||| |||||||| || ||||||||||||||
Db        1021 AGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCACCGAGAAAGTTAAAGA 1080

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               ||||  || |||||||||| |||||||| |||| || |||||||||| ||||||||||||
Db        1081 AATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACACTGGACCCTGGGGCCT 1140

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        1141 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1200

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||||||||||||||||||||||||| ||| |||||||||||||||||| |||||||||
Db        1201 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAAT 1260

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||| ||||| |||||||| |||||||||||||| ||||||||||||||||||| |||
Db        1261 CGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAGTAA 1320

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  ||||||||||||||||||||||| | |||||||||||||||||||||  |
Db        1321 AATCTACCACGGACCCCTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAA 1380

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||||||||||||||||||||||||||| ||||||| | ||||||||| |||
Db        1381 AGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1440

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||  |||||  ||||||||||||||||| | |||||||||||||||||||
Db        1441 AAGCAGTTAGAGTGGTTGTTGGCCAACCTCCCCAACAGCAGTTGGGTTTTCCTGTTGAGA 1500

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               ||| |||||||||||||||| || || |||||||||| |  |||||||||||| ||| ||
Db        1501 GGGGGGACTGAGAGACAGGAATAACTAGATTTCCTAGACCAACTAAGAATCCCTAAGACT 1560

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||||| |||| |||||||||  |||||||||||||||||||| |||
Db        1561 AGCTGGGAAGGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCA 1620

Qy         661 ACCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAA 711
               |||||||         ||||||||||||||||||||||| ||||||||||||||| ||||
Db        1621 ACCAATCAGATACTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGATAAA 1680

Qy         712 GCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAA 771
               | ||||||||||||||| |||||||| |||||||| ||| |||||| ||| |||||||||
Db        1681 GAAATAGCCAATCATCTGTTGCCTGACAGCACAGCAGGAGGGACAATGATCGGGATATAA 1740

Qy         772 ACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAG 831
               || ||||||||| ||||||| ||||| |||| |||||||||||||||| |||||||||||
Db        1741 ACCCAGGCATTCGAGCCAGCTACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAG 1800

Qy         832 CTCTGTTTTCACTCTATT 849
               |||||| |||||||||||
Db        1801 CTCTGTCTTCACTCTATT 1818


RESULT 30
ABA25978
ID   ABA25978 standard; DNA; 1894 BP.
XX
AC   ABA25978;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #4444 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000666.
```

```
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts.
XX
PS   Claim 1; SEQ ID NO 4444; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease. Note: The sequence data for this patent did not
CC   form part of the printed specification, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

  Query Match             53.0%;  Score 704.8;  DB 4;  Length 1894;
  Best Local Similarity   89.7%;  Pred. No. 3e-220;
  Matches  770;  Conservative    0;  Mismatches   79;  Indels    9;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||| |||||||||| |||||||||
Db        961 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGGGAACCTTTTTATTTTT 1020

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              || |||| |||||||||  |||||||||||||| |||||||| || ||||||||||||||
Db       1021 AGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCACCGAGAAAGTTAAAGA 1080

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  || |||||||||| |||||||| |||| || |||||||||| ||||||||||||
Db       1081 AATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACACTGGACCCTGGGGCCT 1140

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1141 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1200

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               |||||||||||||||||||||||||| ||| |||||||||||||||||| |||||||||
Db       1201 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAAT 1260

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||| ||||| |||||||| |||||||||||||| ||||||||||||||||||| |||
Db       1261 CGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAGTAA 1320

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||  ||||||||||||||||||||||| | |||||||||||||||||||||  |
Db       1321 AATCTACCACGGACCCCTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAA 1380

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||||||||||||||||||||||||||| ||||||| | ||||||||| |||
Db       1381 AGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1440

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||  |||||  ||||||||||||||||| | |||||||||||||||||||
Db       1441 AAGCAGTTAGAGTGGTTGTTGGCCAACCTCCCCAACAGCAGTTGGGTTTTCCTGTTGAGA 1500

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||| |||||||||||||||| || || |||||||||| |  |||||||||||| ||| ||
Db       1501 GGGGGGACTGAGAGACAGGAATAACTAGATTTCCTAGACCAACTAAGAATCCCTAAGACT 1560

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||||| |||| |||||||||  |||||||||||||||||||| |||
Db       1561 AGCTGGGAAGGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCA 1620

Qy        661 ACCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAA 711
              |||||||         ||||||||||||||||||||||| ||||||||||||||| ||||
Db       1621 ACCAATCAGATACTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGATAAA 1680

Qy        712 GCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAA 771
              | ||||||||||||||| |||||||| |||||||| ||| |||||| ||| |||||||||
Db       1681 GAAATAGCCAATCATCTGTTGCCTGACAGCACAGCAGGAGGGACAATGATCGGGATATAA 1740

Qy        772 ACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAG 831
              || ||||||||| ||||||| ||||| |||| |||||||||||||||| |||||||||||
Db       1741 ACCCAGGCATTCGAGCCAGCTACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAG 1800

Qy        832 CTCTGTTTTCACTCTATT 849
              |||||| |||||||||||
Db       1801 CTCTGTCTTCACTCTATT 1818

Search completed: January 22, 2005, 16:40:42
Job time : 618.913 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 12:03:14 ; Search time 123.063 Seconds
                                             (without alignments)
                                             7676.074 Million cell updates/sec

Title:          US-09-319-156B-12
Perfect score:  1329
Sequence:       1 tcaaaatcgaagagctttag...........gcaaggatcccccagtaaca 1329

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters:      1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                      %
Result              Query
  No.     Score    Match Length DB  ID                           Description
 ------------------------------------------------------------------------------
     1     1326    99.8    1329  4  US-08-979-847B-108            Sequence 108, App
     2    731.8    55.1    2946  3  US-09-175-928-3               Sequence 3, Appli
     3      545    41.0     635  4  US-08-979-847B-102            Sequence 102, App
     4    344.6    25.9     443  4  US-09-513-999C-21394          Sequence 21394, A
     5      302    22.7    1481  4  US-08-979-847B-105            Sequence 105, App
     6    265.2    20.0     800  4  US-08-979-847B-127            Sequence 127, App
     7    257.8    19.4     764  4  US-08-979-847B-126            Sequence 126, App
     8      208    15.7     279  1  US-08-686-878A-50             Sequence 50, Appl
     9      208    15.7     279  1  US-08-721-489-4               Sequence 4, Appli
 c  10    194.6    14.6     456  4  US-09-621-976-9366            Sequence 9366, Ap
    11      166    12.5    1722  3  US-08-691-563C-58             Sequence 58, Appl
    12      166    12.5    1722  4  US-09-374-766-58              Sequence 58, Appl
    13      166    12.5    1722  4  US-08-979-847B-54             Sequence 54, Appl
    14      165    12.4    1859  3  US-08-691-563C-46             Sequence 46, Appl
    15      165    12.4    1859  4  US-09-374-766-46              Sequence 46, Appl
    16      165    12.4    1859  4  US-08-979-847B-42             Sequence 42, Appl
    17    161.8    12.2   80246  3  US-09-078-294-4               Sequence 4, Appli
    18    151.8    11.4   80595  3  US-09-078-294-3               Sequence 3, Appli
 c  19    123.6     9.3     506  4  US-09-621-976-2599            Sequence 2599, Ap
    20    107.8     8.1    3910  3  US-09-120-653D-1              Sequence 1, Appli
    21    103.2     7.8     149  4  US-09-513-999C-24228          Sequence 24228, A
    22     85.8     6.5    2908  4  US-09-799-451-249             Sequence 249, App
    23     80.6     6.1    1704  1  US-08-007-282B-1              Sequence 1, Appli
    24     80.6     6.1    2518  3  US-09-011-745-1               Sequence 1, Appli
    25     80.6     6.1    4776  4  US-09-555-352-1               Sequence 1, Appli
    26     80.6     6.1    5865  3  US-09-011-745-8               Sequence 8, Appli
    27     78.2     5.9     464  3  US-09-120-653D-6              Sequence 6, Appli
```

```
    28       76     5.7     492  4  US-09-621-976-11057           Sequence 11057, A
    29     73.4     5.5     442  4  US-09-621-976-17982           Sequence 17982, A
    30     69.6     5.2    2809  4  US-09-620-312D-171            Sequence 171, App
    31       68     5.1    2342  4  US-09-904-615-12              Sequence 12, Appl
    32       57     4.3    7218  1  US-08-232-463-14              Sequence 14, Appl
    33     50.8     3.8    8060  3  US-08-766-528-1               Sequence 1, Appli
    34     50.8     3.8    8060  4  US-09-661-858-1               Sequence 1, Appli
    35     49.2     3.7    8132  3  US-08-766-528-3               Sequence 3, Appli
    36     49.2     3.7    8132  4  US-09-661-858-3               Sequence 3, Appli
    37     48.4     3.6     304  4  US-09-513-999C-26710          Sequence 26710, A
    38     47.8     3.6     502  4  US-09-621-976-9610            Sequence 9610, Ap
    39     46.4     3.5    1911  1  US-08-258-420-8               Sequence 8, Appli
    40     46.4     3.5    1965  1  US-08-258-420-9               Sequence 9, Appli
    41     46.4     3.5    3925  3  US-09-011-745-9               Sequence 9, Appli
 c  42     46.4     3.5    8202  1  US-08-258-420-13              Sequence 13, Appl
    43     45.2     3.4     342  4  US-09-513-999C-13878          Sequence 13878, A
    44     44.8     3.4    6028  3  US-09-011-745-5               Sequence 5, Appli
    45     44.6     3.4    1989  3  US-09-232-278A-5              Sequence 5, Appli

                              ALIGNMENTS


RESULT 1
US-08-979-847B-108
; Sequence 108, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,847B
;            FILING DATE: 26-No. 6582703-1997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;  INFORMATION FOR SEQ ID NO: 108:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1329 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 108:
US-08-979-847B-108
```

```
  Query Match             99.8%;  Score 1326;  DB 4;  Length 1329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1329;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260

Qy       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

Qy       1321 CCAGTAACA 1329
              |||||||||
Db       1321 CCAGTAACA 1329


RESULT 2
US-09-175-928-3
; Sequence 3, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT: McCoy, John M.
;  APPLICANT: LaVallie, Edward R.
;  APPLICANT: Collins-Racie, Lisa A.
;  APPLICANT: Evans, Cheryl
;  APPLICANT: Merberg, David
;  APPLICANT: Treacy, Maurice
;  APPLICANT: Mi, Sha
;  APPLICANT: Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6006B.AJ172A
;  CURRENT APPLICATION NUMBER: US/09/175,928A
;  CURRENT FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-175-928-3

  Query Match             55.1%;  Score 731.8;  DB 3;  Length 2946;
  Best Local Similarity   90.7%;  Pred. No. 7.7e-240;
  Matches  794;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       2067 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 2126

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db       2127 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2186

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db       2187 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2246

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2247 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2306

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||  ||| |||||||||| |||||||||||||||||
Db       2307 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2366

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       2367 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2414

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2415 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2474

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2475 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2534

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2535 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2594

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               ||  ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       2595 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2654

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2655 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2714

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2715 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2774

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2775 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2834

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2835 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2894

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2895 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2929


RESULT 3
US-08-979-847B-102
; Sequence 102, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
;          CURRENT APPLICATION DATA:
;               APPLICATION NUMBER: US/08/979,847B
;               FILING DATE: 26-No. 6582703-1997
;               CLASSIFICATION: <Unknown>
;          ATTORNEY/AGENT INFORMATION:
;               NAME: BERRIDGE, WILLIAM P.
;               REGISTRATION NUMBER: 30,024
;               REFERENCE/DOCKET NUMBER: WPB 39046A
;          TELECOMMUNICATION INFORMATION:
;               TELEPHONE: 703-836-6400
;               TELEFAX: 703-836-2787
;     INFORMATION FOR SEQ ID NO: 102:
;          SEQUENCE CHARACTERISTICS:
;               LENGTH: 635 base pairs
;               TYPE: nucleic acid
;               STRANDEDNESS: single
;               TOPOLOGY: linear
;          MOLECULE TYPE: cDNA
;          SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-08-979-847B-102

  Query Match             41.0%;  Score 545;  DB 4;  Length 635;
  Best Local Similarity   92.4%;  Pred. No. 3.1e-176;
  Matches  572;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
              ||||||||||| ||| ||||||||||||||||||||||||||||  || ||||||| |||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
              ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436
              ||||||||||||||||| | |||||| ||||||||||||||||||| |||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496
              |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616
              |||||||||||||||||||||||  || ||||||||| ||||||| |||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676
              ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736
              |||||||||||||| ||||||||| ||||||||||| |||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796
              |||||||||||||| |||||||||| ||||||||||| ||||||||| |||| ||||| |
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856
              ||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy        857 ATTAAATCATGCAACTGCA 875
              |||||||| |||||||| |
Db        601 ATTAAATCTTGCAACTGAA 619


RESULT 4
US-09-513-999C-21394
; Sequence 21394, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 21394
;   LENGTH: 443
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 102
;   OTHER INFORMATION: k=g or t
US-09-513-999C-21394

  Query Match             25.9%;  Score 344.6;  DB 4;  Length 443;
  Best Local Similarity   88.9%;  Pred. No. 1.3e-107;
  Matches  394;  Conservative    1;  Mismatches   45;  Indels    3;  Gaps    2;

Qy        766 ATATAAACTCAGGCATTCAAGCCAGCAACAGCA-ACCCCCTTTGGGTCCCCTCCCATTGT 824
              |||||||| ||||||||| |||  |||||||||  ||||||||||||||| |||| ||||
Db          3 ATATAAACCCAGGCATTCGAGCTGGCAACAGCAGCCCCCCTTTGGGTCCCTTCCCTTTGT 62

Qy        825 ATGGGAGCTCTGTTTTCACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGG 884
              |||||||  |||||||||  |||||||||||||||||||| :||||||||||||||||||
Db         63 ATGGGAG--CTGTTTTCATGCTATTTCACTCTATTAAATCTKGCAACTGCACTCTTCTGG 120

Qy        885 TCCGTGTTTTTTATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCA 944
              ||| ||||| ||| ||||| ||||||||||||| || || ||||||||||||||||||||
Db        121 TCCATGTTTCTTACGGCTCGAGCTGAGCTTTTGCTCACCGTCCACCACTGCTGTTTGCCA 180

Qy        945 CCGTCACAGACCCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCT 1004
              ||  | |||||| || ||||||| |||||||| ||||||| |||| |||||| |||||||
Db        181 CCACCGCAGACCTGCCGCTGACTCCCATCCCTCTGGATCCTGCAGGGTGTCCGCTGTGCT 240

Qy       1005 CCTGATCCAGCGAGGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCC 1064
              |||||||||||||||  ||||||||| ||||| ||  |||||||||||||||||||||||
Db        241 CCTGATCCAGCGAGGCGCCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTTCC 300

Qy       1065 TGCATGGCTAAGTGCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCAT 1124
              |||| |||||||||||||||||||| ||||| || ||||||||| |||||||||||||||
Db        301 TGCACGGCTAAGTGCCTGGGTTTGTTCTAATTGAGCTGAACACTAGTCACTGGGTTCCAT 360

Qy       1125 GGTTCTCTTCCATGACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAG 1184
              ||||||||||  ||||||||||||||||||||| |||||||||| ||| |||||||||||
Db        361 GGTTCTCTTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACCACATGGCCCAAG 420

Qy       1185 ATTCCATTCCTTGGTATCTGTGA 1207
              |||||||||||||| ||| ||||
Db        421 ATTCCATTCCTTGGAATCCGTGA 443


RESULT 5
US-08-979-847B-105
; Sequence 105, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
```

```
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLA
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 105:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1481 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 105:
US-08-979-847B-105

  Query Match             22.7%;  Score 302;  DB 4;  Length 1481;
  Best Local Similarity   92.7%;  Pred. No. 1.2e-92;
  Matches  317;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGAACCTGTTTATTTTT 1199

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||| |||||| ||||||| ||||||||||||||  ||||  | |||||||||||||||||
Db       1200 AGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGA 1259

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| |||||||||||||||| ||||||||||| | || |||||||||
Db       1260 AATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCT 1319

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||  |||||||||||||||||||||||||||| |||||||
Db       1320 CCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATT 1379

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               |||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||
Db       1380 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAAT 1439

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCA 342
              |||||||||||||||||| || ||  | ||||||||||||||
Db       1440 TGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481


RESULT 6
US-08-979-847B-127
; Sequence 127, Application US/08979847B
; Patent No. 6582703
;   GENERAL INFORMATION:
```

```
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 127:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 800 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 127:
US-08-979-847B-127

  Query Match             20.0%;  Score 265.2;  DB 4;  Length 800;
  Best Local Similarity   88.2%;  Pred. No. 3.3e-80;
  Matches  299;  Conservative    0;  Mismatches   39;  Indels    1;  Gaps    1;

Qy        992 TGTCCACTGTGCTCCTGATCCAGC-GAGGTACCCATTGCCACTCCCGATCAGGCTAAAGG 1050
              ||||| ||||||||||||||||||  |||  ||||||||| ||||| ||  |||||||||
Db          1 TGTCCGCTGTGCTCCTGATCCAGCACAGGCGCCCATTGCCTCTCCCAATTGGGCTAAAGG 60

Qy       1051 CTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTCCTAATAGAACTGAACACTGG 1110
              ||||||||||||||||||  ||||||||||||||||  ||||||| || ||||||||| |
Db         61 CTTGCCATTGTTCCTGCACAGCTAAGTGCCTGGGTTCATCCTAATCGAGCTGAACACTAG 120

Qy       1111 TCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCTTCTAATAGAGCTATAACACTCA 1170
              ||||||||||||| ||||||||||||||||||| ||||||||||||||||||||||||||
Db        121 TCACTGGGTTCCACGGTTCTCTTCCATGACCCATGGCTTCTAATAGAGCTATAACACTCA 180

Qy       1171 CCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGA 1230
              | |||||| ||||||||||||||||||| ||| ||||| |||||||||||||||||||||
Db        181 CTGCATGGTCCAAGATTCCATTCCTTGGAATCCGTGAGACCAAGAACCCCAGGTCAGAGA 240

Qy       1231 ANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGCCATTTTGGTAGCGGCCCACCAC 1290
              |    |||||||||||||| | |||||  ||||||  ||||||||| |||||||| ||||
Db        241 ACACAAGGCTTGCCACCATGTTGGAAGCAGCCCACCACCATTTTGGAAGCGGCCCGCCAC 300

Qy       1291 CATCTTGGGAGCTGTGGGAGCAAGGATCCCCCAGTAACA 1329
               |||||||||||| |||||||||||| ||||  ||||||
Db        301 TATCTTGGGAGCTCTGGGAGCAAGGACCCCCAGGTAACA 339
```

```
RESULT 7
US-08-979-847B-126
; Sequence 126, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;        APPLICANT: PERRON, HERVE
;                   BESEME, FREDERIC
;                   BEDIN, FREDERIC
;                   PARANHOS-BACCALA, GLAUCIA
;                   KOMURIAN-PRADEL, FLORENCE
;                   JOLIVET-REYNAUD, COLETTE
;                   MANDRAND, BERNARD
;                   GARSON, JEREMY
;                   TUKE, PHILIP
;        TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                            ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                            THERAPEUTIC PURPOSES
;        NUMBER OF SEQUENCES: 210
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: OLIFF & BERRIDGE, PLC
;             STREET: P.O. BOX 19928
;             CITY: ALEXANDRIA
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22320
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/979,847B
;             FILING DATE: 26-No. 6582703-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BERRIDGE, WILLIAM P.
;             REGISTRATION NUMBER: 30,024
;             REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-836-6400
;             TELEFAX: 703-836-2787
;   INFORMATION FOR SEQ ID NO: 126:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 764 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 126:
US-08-979-847B-126

  Query Match             19.4%;  Score 257.8;  DB 4;  Length 764;
  Best Local Similarity   88.2%;  Pred. No. 1.1e-77;
  Matches  291;  Conservative    0;  Mismatches   38;  Indels    1;  Gaps    1;

Qy        992 TGTCCACTGTGCTCCTGATCCAGC-GAGGTACCCATTGCCACTCCCGATCAGGCTAAAGG 1050
              ||||| ||||||||||||||||||  |||  ||||||||| ||||| ||  |||||||||
Db          1 TGTCCGCTGTGCTCCTGATCCAGCACAGGCGCCCATTGCCTCTCCCAATTGGGCTAAAGG 60

Qy       1051 CTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTCCTAATAGAACTGAACACTGG 1110
              ||||||||||||||||||  ||||||||||||||||  ||||||| || ||||||||| |
Db         61 CTTGCCATTGTTCCTGCACAGCTAAGTGCCTGGGTTCATCCTAATCGAGCTGAACACTAG 120

Qy       1111 TCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCTTCTAATAGAGCTATAACACTCA 1170
              ||||||||||||| ||||||||||||||||||| ||||||||||||||||||||||||||
Db        121 TCACTGGGTTCCACGGTTCTCTTCCATGACCCATGGCTTCTAATAGAGCTATAACACTCA 180

Qy       1171 CCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGA 1230
              | |||||| ||||||||||||||||||| ||| ||||| |||||||||||||||||||||
Db        181 CTGCATGGTCCAAGATTCCATTCCTTGGAATCCGTGAGACCAAGAACCCCAGGTCAGAGA 240
```

```
Qy       1231 ANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGCCATTTTGGTAGCGGCCCACCAC 1290
              |    |||||||||||||| | |||||  ||||||  ||||||||| ||| |||| ||||
Db        241 ACACAAGGCTTGCCACCATGTTGGAAGCAGCCCACCACCATTTTGGAAGCAGCCCGCCAC 300

Qy       1291 CATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
               |||||||||||| |||||||||||| |||
Db        301 TATCTTGGGAGCTCTGGGAGCAAGGACCCC 330

RESULT 8
US-08-686-878A-50
; Sequence 50, Application US/08686878A
; Patent No. 5708157
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Evans, Cheryl
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/686,878A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  50:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  279 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-686-878A-50

  Query Match             15.7%;  Score 208;  DB 1;  Length 279;
  Best Local Similarity   85.1%;  Pred. No. 6.5e-61;
  Matches  223;  Conservative    1;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        614 ACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAG 673
              :|| |||||| ||||||||| |||| ||||||   || | ||||  ||||||||||||||
Db          1 RCCACATCCACCTTTAAACACGGGGNTTGCAAANAAGATNACACTTGACCAATCAGAGAG 60

Qy        674 CTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGC 733
               ||| ||||||| | ||  |||||||||||||||||||| ||||||||||||||||||||
Db         61 NTCANTAAAATGATNATTNGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGC 120

Qy        734 CTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAA 793
              ||||||||||||| ||| |||||| ||| ||||||||||| || |  ||  |||| ||||
Db        121 CTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTNGAGCCGGCAA 180

Qy        794 CAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCAC 853
```

```
              | |||||||||||||||||||||||| ||||||||||||| ||||||||  ||||||||
Db        181 CGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTNTGTTTTCATGCTATTTCAN 240

Qy        854 TCTATTAAATCATGCAACTGCA 875
              | ||||||||  ||||||||||
Db        241 TNTATTAAATNTTGCAACTGCA 262

RESULT 9
US-08-721-489-4
; Sequence 4, Application US/08721489
; Patent No. 5786465
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/721,489
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  279 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-721-489-4

  Query Match             15.7%;  Score 208;  DB 1;  Length 279;
  Best Local Similarity   85.1%;  Pred. No. 6.5e-61;
  Matches  223;  Conservative    1;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        614 ACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAG 673
              :|| |||||| ||||||||| |||| ||||||   || | ||||  ||||||||||||||
Db          1 RCCACATCCACCTTTAAACACGGGGNTTGCAAANAAGATNACACTTGACCAATCAGAGAG 60

Qy        674 CTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGC 733
               ||| ||||||| | ||  |||||||||||||||||||| ||||||||||||||||||||
Db         61 NTCANTAAAATGATNATTNGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGC 120

Qy        734 CTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAA 793
              ||||||||||||| ||| |||||| ||| ||||||||||| || |  ||  |||| ||||
Db        121 CTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTTTTNGAGCCGGCAA 180

Qy        794 CAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCAC 853
              | |||||||||||||||||||||||| ||||||||||||| ||||||||  ||||||||
Db        181 CGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTNTGTTTTCATGCTATTTCAN 240
```

```
Qy        854 TCTATTAAATCATGCAACTGCA 875
              | ||||||||  ||||||||||
Db        241 TNTATTAAATNTTGCAACTGCA 262

RESULT 10
US-09-621-976-9366/c
; Sequence 9366, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 9366
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-9366

  Query Match             14.6%;  Score 194.6;  DB 4;  Length 456;
  Best Local Similarity   84.2%;  Pred. No. 3.6e-56;
  Matches  262;  Conservative    0;  Mismatches   26;  Indels   23;  Gaps    3;

Qy        548 CTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGG 607
              |||||||||||||||||||||||||||||||| ||||||||||||| ||||||| |||||
Db        307 CTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGG 248

Qy        608 AAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC 667
              ||||||||| |||||| |||||| || |||||||||||||||||||||||||||||| ||
Db        247 AAGGTGACCACATCCACCTTTAAGCAGGGGGCTTGCAACTTAGCTCACACCCGACCAGTC 188

Qy        668 ---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAG 718
                       || |||||||||||||||||||| |||| ||||||||||||||| ||||||
Db        187 AGGTAGTAAAGCGAGCTCACTAAAATGCTAATTAGGCTAAAACAGGAGGTAAA-CAATAG 129

Qy        719 CCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGG 778
              ||||||||||||  |||  |||||||||| ||| |||||| ||| |||||||||||
Db        128 CCAATCATCTATCTCCTCAGAGCACAGCAGGAGGGACAATGATAGGGATATAAAC----- 74

Qy        779 CATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTT 838
                      |  ||||| ||  ||| || ||||||||||||| ||||||||||||||||||
Db         73 --------CTGGCAACGGCTGCCCTCTCTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTT 22

Qy        839 TTCACTCTATT 849
              |||||||||||
Db         21 TTCACTCTATT 11

RESULT 11
US-08-691-563C-58
; Sequence 58, Application US/08691563C
; Patent No. 6001987
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563C
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 base pairs
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-691-563C-58

  Query Match             12.5%;  Score 166;  DB 3;  Length 1722;
  Best Local Similarity   60.5%;  Pred. No. 5.9e-46;
  Matches  333;  Conservative    0;  Mismatches  200;  Indels   17;  Gaps    3;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              ||||| ||  |||    | || | || || ||  | | ||| |||  |||    || |||
Db        547 CAAAACCGCCGAGGCCCACACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTA 606

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
              |||||||| || ||||  ||   ||| || || || ||  |    ||           |
Db        607 GGGGAAGAGTGTTGTTTTTACACTAACCAGTCAGGGATAGTACGAGATGCCACCTGGCAT 666

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |    ||  |    |    ||    | |   ||||  |||| ||    ||        |
Db        667 TTACAGGAAAGGGCTTCTGATATCAGACAATGCCTTTCAAACTCTTATACCAA------C 720

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||| |    |||    | |||  ||| || ||  ||||  ||  | | |||||  ||  |  |
Db        721 CTCTGGAGTTGGGCAACATGGCTTCTTCCATTTCTAGGTCCCATGGCAGCCATCTTGCTG 780

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
              ||||||  |||||| |||||||| || || || |  ||||| || ||||| || || || ||
Db        781 TTACTCACCTTTGGGCCCTGTATTTTTAAGCTTCTTGTCAAATTTGTTTCCTCTAGGATC 840

Qy        302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
              |||||  | |||||||| || ||  | |||||||||||||| |||   || |  |  ||
Db        841 GAAGCCATCAAGCTACAGATGGTCTTACAAATGGAACCCCAAATGAGTTCAACTAACAAC 900

Qy        362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
               ||||||  ||||||||||| ||  |  || |    | | ||          |   |  |
Db        901 TTCTACCAAGGACCCCTGGAACGATCCACTGGCACTTCCACT----------AGCCTAGA 950

Qy        422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
              |   ||||||  || || |   ||||||||   ||||| ||   | || || ||| ||||
Db        951 GATTCCCCTCTGGAAGACACTACAACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGA 1010

Qy        482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
              || || |||||| ||  || |||||  | |||||||| | ||||| | ||||||| ||||
Db       1011 AGTAGCTAGAGCGGTCATCGGCCAA-ATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAG 1069

Qy        542 GGTGGACTGA 551
              || ||| |||
Db       1070 GGGGGATTGA 1079
```

```
RESULT 12
US-09-374-766-58
; Sequence 58, Application US/09374766
; Patent No. 6579526
;  GENERAL INFORMATION:
;    APPLICANT:  Herve PERRON
;    APPLICANT:  Frederic BESEME
;    APPLICANT:  Frederic BEDIN
;    APPLICANT:  Glaucia PARANHOS-BACCALA
;    APPLICANT:  Florence KOMURIAN-PRADEL
;    APPLICANT:  Colette JOLIVET
;    APPLICANT:  Bernard MANDRAND
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/374,766
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/691,563
;      FILING DATE:  02-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berridge, William P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 38588
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 base pairs
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-374-766-58

  Query Match             12.5%;  Score 166;  DB 4;  Length 1722;
  Best Local Similarity   60.5%;  Pred. No. 5.9e-46;
  Matches  333;  Conservative    0;  Mismatches  200;  Indels   17;  Gaps    3;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              ||||| ||  |||    | || | || || ||  | | ||| |||  |||    || |||
Db        547 CAAAACCGCCGAGGCCCACACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTA 606

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
              |||||||| || ||||  ||   ||| || || || ||  |    ||           |
Db        607 GGGGAAGAGTGTTGTTTTACACTAACCAGTCAGGGATAGTACGAGATGCCACCTGGCAT 666

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |    ||  |    |    ||    | |   ||||  |||| ||    ||        |
Db        667 TTACAGGAAAGGGCTTCTGATATCAGACAATGCCTTTCAAACTCTTATACCAA------C 720

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||| |    |||    | |||  ||| || ||  |||| ||  | ||||| ||  |  |
Db        721 CTCTGGAGTTGGGCAACATGGCTTCTTCCATTTCTAGGTCCCATGGCAGCCATCTTGCTG 780

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
              ||||||  |||||| |||||||| || ||  | ||||| || ||||| || || || ||
Db        781 TTACTCACCTTTGGGCCCTGTATTTTTAAGCTTCTTGTCAAATTTGTTTCCTCTAGGATC 840

Qy        302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
              |||||  | |||||||| || ||  | |||||||||||||| |||   || |  |  ||
Db        841 GAAGCCATCAAGCTACAGATGGTCTTACAAATGGAACCCCAAATGAGTTCAACTAACAAC 900

Qy        362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
               ||||||  ||||||||||| ||  |  || |    | | ||          |   |  |
Db        901 TTCTACCAAGGACCCCTGGAACGATCCACTGGCACTTCCACT----------AGCCTAGA 950

Qy        422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
              |   ||||||  || || |   ||||||||   ||||| ||   | || || ||| ||||
Db        951 GATTCCCCTCTGGAAGACACTACAACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGA 1010

Qy        482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
              || || |||||| ||  || |||||  | |||||||| | ||||| | ||||||| ||||
Db       1011 AGTAGCTAGAGCGGTCATCGGCCAA-ATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAG 1069

Qy        542 GGTGGACTGA 551
              || ||| |||
Db       1070 GGGGGATTGA 1079

RESULT 13
US-08-979-847B-54
; Sequence 54, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLA
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,847B
;            FILING DATE: 26-No. 6582703-1997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;  INFORMATION FOR SEQ ID NO: 54:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1722 base pairs
;            TYPE: nucleotide
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-08-979-847B-54
```

```
  Query Match             12.5%;  Score 166;  DB 4;  Length 1722;
  Best Local Similarity   60.5%;  Pred. No. 5.9e-46;
  Matches  333;  Conservative    0;  Mismatches  200;  Indels   17;  Gaps    3;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              ||||| ||  |||    | || | || || ||  | | ||| |||  |||    || |||
Db        547 CAAAACCGCCGAGGCCCACACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTA 606

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
              |||||||| || ||||  ||   ||| || || || ||  |    ||           |
Db        607 GGGGAAGAGTGTTGTTTTTACACTAACCAGTCAGGGATAGTACGAGATGCCACCTGGCAT 666

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |    ||  |    |    ||    | |   ||||  |||| ||    ||        |
Db        667 TTACAGGAAAGGGCTTCTGATATCAGACAATGCCTTTCAAACTCTTATACCAA------C 720

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||| |    |||    | |||  ||| || ||  |||| ||  | ||||| ||  |  |
Db        721 CTCTGGAGTTGGGCAACATGGCTTCTTCCATTTCTAGGTCCCATGGCAGCCATCTTGCTG 780

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
              ||||||  |||||| |||||||| || ||  | ||||| || ||||| || || || ||
Db        781 TTACTCACCTTTGGGCCCTGTATTTTTAAGCTTCTTGTCAAATTTGTTTCCTCTAGGATC 840

Qy        302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
              |||||  | |||||||| || ||  | |||||||||||||| |||   || |  |  ||
Db        841 GAAGCCATCAAGCTACAGATGGTCTTACAAATGGAACCCCAAATGAGTTCAACTAACAAC 900

Qy        362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
               ||||||  ||||||||||| ||  |  || |  | | | ||          |   |  |
Db        901 TTCTACCAAGGACCCCTGGAACGATCCACTGGCACTTCCACT----------AGCCTAGA 950

Qy        422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
              |   ||||||  || || |   ||||||||   ||||| ||   | || || ||| ||||
Db        951 GATTCCCCTCTGGAAGACACTACAACTGCAGGGCCCCTTCTTTGCCCCTATCCAGCAGGA 1010

Qy        482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
              || || |||||| ||  || |||||  | |||||||| | ||||| | ||||||| ||||
Db       1011 AGTAGCTAGAGCGGTCATCGGCCAA-ATTCCCAACAGCAGTTGGGGTGTCCTGTTTAGAG 1069

Qy        542 GGTGGACTGA 551
              || ||| |||
Db       1070 GGGGGATTGA 1079


RESULT 14
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
;  GENERAL INFORMATION:
;     APPLICANT:  Herve PERRON
;     APPLICANT:  Frederic BESEME
;     APPLICANT:  Frederic BEDIN
;     APPLICANT:  Glaucia PARANHOS-BACCALA
;     APPLICANT:  Florence KOMURIAN-PRADEL
;     APPLICANT:  Colette JOLIVET
;     APPLICANT:  Bernard MANDRAND
;     TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;     TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;     TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;     NUMBER OF SEQUENCES:  92
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Oliff & Berridge
;       STREET:  700 South Washington Street, Suite 300
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/691,563C
;       FILING DATE:  02-AUG-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berridge, William P.
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 38588
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703-836-6400
;       TELEFAX:  703-836-2787
;   INFORMATION FOR SEQ ID NO:  46:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1859 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-691-563C-46

  Query Match             12.4%;  Score 165;  DB 3;  Length 1859;
  Best Local Similarity   58.3%;  Pred. No. 1.4e-45;
  Matches  369;  Conservative    0;  Mismatches  247;  Indels   17;  Gaps    4;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              ||||| ||   ||   ||||| | || || ||  | | ||| |||  |||    || |||
Db       1197 CAAAACCGTCAAGGCCTAGACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTA 1256

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
               ||||||| || |||   ||   ||| || || || ||  |   |||||   |    |
Db       1257 AGGGAAGAGTGTTGTCTTTACACTAACCAGTCAGGGATAGTA--TGAGATGCTGCCCGGC 1314

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
              ||||       ||        | | ||||  ||       | | |    |||    |  |
Db       1315 ATTTACAGAAAAAGGCTTCTGAAATCAGACAACGCCTTTCAAATT----CCTATACCAAC 1370

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||| |    |||    | |||  ||| |||||  || | ||  | || || ||  |  |
Db       1371 CTCTGGAGTTGGGCAACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTA 1430

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
              ||||||  |||||| |||||||| || ||| ||||||| || ||||| ||||| || ||
Db       1431 TTACTCGCCTTTGGGCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATC 1490

Qy        302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
              || ||  | |||||||| || ||  | |||||||||||||| |||   || |  |  ||
Db       1491 GAGGCCATCAAGCTACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAAC 1550

Qy        362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
               ||||| | |||||||| ||||  ||  || | |  | | |||   | ||
Db       1551 TTCTACTGAGGACCCCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTT----- 1605

Qy        422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
                  ||||||  ||||| |   | ||||||   ||||  ||   | || || ||| ||||
Db       1606 ----CCCCTCTGGAGGACACTACCACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGA 1661

Qy        482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
              || || |||||||||  |   ||||  | ||||| || |  |||| | ||| ||| ||||
Db       1662 AGTAGCTAGAGCAGTCATTGCCCAA--TTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAG 1719

Qy        542 GGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTA 601
               | ||| ||||||      | ||||||| |||   | | |     || |     | | |
Db       1720 TGGGGATTGAGAGGTGAAGCCAGCTGGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTT 1779

Qy        602 NCTGGGAAGGTGACCGCATCCATCTTTAAACAT 634
                ||   || | |  |  |  |  |  ||  ||
Db       1780 TGTGTCTAGCTAAAGGATTGTAAATGCAACAAT 1812


RESULT 15
US-09-374-766-46
; Sequence 46, Application US/09374766
; Patent No. 6579526
;   GENERAL INFORMATION:
;     APPLICANT:  Herve PERRON
;     APPLICANT:  Frederic BESEME
;     APPLICANT:  Frederic BEDIN
;     APPLICANT:  Glaucia PARANHOS-BACCALA
;     APPLICANT:  Florence KOMURIAN-PRADEL
;     APPLICANT:  Colette JOLIVET
;     APPLICANT:  Bernard MANDRAND
;     TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;     TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;     TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;     NUMBER OF SEQUENCES:  92
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Oliff & Berridge
;       STREET:  700 South Washington Street, Suite 300
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22314
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/374,766
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/691,563
;       FILING DATE:  02-AUG-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berridge, William P.
;       REGISTRATION NUMBER:  30,024
;       REFERENCE/DOCKET NUMBER:  WPB 38588
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703-836-6400
;       TELEFAX:  703-836-2787
;   INFORMATION FOR SEQ ID NO:  46:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1859 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-09-374-766-46

  Query Match             12.4%;  Score 165;  DB 4;  Length 1859;
  Best Local Similarity   58.3%;  Pred. No. 1.4e-45;
  Matches  369;  Conservative    0;  Mismatches  247;  Indels   17;  Gaps    4;

Qy             2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
                 ||||| ||   ||   ||||| | || || ||  | | ||| |||  |||    || |||
Db          1197 CAAAACCGTCAAGGCCTAGACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTA 1256

Qy            62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
                  ||||||| || |||   ||   ||| || || || ||  |   |||||   |    |
Db          1257 AGGGAAGAGTGTTGTCTTTACACTAACCAGTCAGGGATAGTA--TGAGATGCTGCCCGGC 1314

Qy           122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
                 ||||       ||        | | ||||  ||       | | |    |||    |  |
Db          1315 ATTTACAGAAAAAGGCTTCTGAAATCAGACAACGCCTTTCAAATT----CCTATACCAAC 1370

Qy           182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
                 ||| |    |||    | |||  ||| |||||  || | ||  | || || ||  |  |
Db          1371 CTCTGGAGTTGGGCAACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTA 1430

Qy           242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
                 ||||||  |||||| |||||||| || ||| ||||||| || ||||| ||||| || ||
Db          1431 TTACTCGCCTTTGGGCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATC 1490

Qy           302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
                 || ||  | |||||||| || ||  | |||||||||||||| |||   || |  |  ||
Db          1491 GAGGCCATCAAGCTACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAAC 1550
```

```
Qy           362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
                  ||||| | |||||||| ||||  ||  || | |  | | |||   | ||
Db          1551 TTCTACTGAGGACCCCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTT----- 1605

Qy           422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
                     ||||||  ||||| |   | ||||||   ||||  ||   | || || ||| ||||
Db          1606 ----CCCCTCTGGAGGACACTACCACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGA 1661

Qy           482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
                 || || |||||||||  |   ||||  | ||||| || |  |||| | ||| ||| ||||
Db          1662 AGTAGCTAGAGCAGTCATTGCCCAA--TTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAG 1719

Qy           542 GGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTA 601
                  | ||| ||||||      | ||||||| |||   | | |     || |     | | |
Db          1720 TGGGGATTGAGAGGTGAAGCCAGCTGGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTT 1779

Qy           602 NCTGGGAAGGTGACCGCATCCATCTTTAAACAT 634
                   ||   || | |  |  |  |  |  ||  ||
Db          1780 TGTGTCTAGCTAAAGGATTGTAAATGCAACAAT 1812

RESULT 16
US-08-979-847B-42
; Sequence 42, Application US/08979847B
; Patent No. 6582703
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL/
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/979,847B
;            FILING DATE: 26-No. 6582703-1997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;    INFORMATION FOR SEQ ID NO: 42:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1859 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-08-979-847B-42
```

```
  Query Match             12.4%;  Score 165;  DB 4;  Length 1859;
  Best Local Similarity   58.3%;  Pred. No. 1.4e-45;
  Matches  369;  Conservative    0;  Mismatches  247;  Indels   17;  Gaps    4;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              ||||| ||   ||   ||||| | || || ||  | | ||| |||  |||    || |||
Db       1197 CAAAACCGTCAAGGCCTAGACCTCCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTA 1256

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
               ||||||| || |||   ||   ||| || || || ||  |   |||||   |    |
Db       1257 AGGGAAGAGTGTTGTCTTTACACTAACCAGTCAGGGATAGTA--TGAGATGCTGCCCGGC 1314

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
              ||||       ||        | | ||||  ||       | | |    |||    |  |
Db       1315 ATTTACAGAAAAAGGCTTCTGAAATCAGACAACGCCTTTCAAATT----CCTATACCAAC 1370

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||| |    |||    | |||  |||||||||  || | ||  | || || ||  |  |
Db       1371 CTCTGGAGTTGGGCAACATGGTTTCTTCCCTTTCTATGTCCCATGGCTGCCATCTTGCTA 1430

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
              ||||||  |||||| |||||||| || ||| ||||||| || ||||| ||||| || ||
Db       1431 TTACTCGCCTTTGGGCCCTGTATTTTTAACCTCCTTGTCAAATTTGTTTCTTCTAGGATC 1490

Qy        302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
              || ||  | |||||||| || ||  | |||||||||||||| |||   || |  |  ||
Db       1491 GAGGCCATCAAGCTACAGATGGTCTTACAAATGGAACCCCAAATGAGCTCAACTATCAAC 1550

Qy        362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
               ||||| | |||||||||||||  ||  || | |  | | |||   | ||
Db       1551 TTCTACTGAGGACCCCTAGACCAACCCCCTGGCCCTTTCACTGGCCTAAAGAGTT----- 1605

Qy        422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
                  ||||||  ||||| |   | ||||||   ||||  ||   | || || ||| ||||
Db       1606 ----CCCCTCTGGAGGACACTACCACTGCAGGGCCCCATCTTTGCCCCTATCCAGAAGGA 1661

Qy        482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
              || || |||||||||  |   ||||  | ||||| || |  |||| | ||| ||| ||||
Db       1662 AGTAGCTAGAGCAGTCATTGCCCAA--TTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAG 1719

Qy        542 GGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTA 601
               | ||| ||||||      | ||||||| |||   |   |     || |     | | |
Db       1720 TGGGGATTGAGAGGTGAAGCCAGCTGGACTTCTGGGTCGGGTGGGGACTTGGAGAACTTT 1779

Qy        602 NCTGGGAAGGTGACCGCATCCATCTTTAAACAT 634
                ||   || | |  |  |  |  |  ||  ||
Db       1780 TGTGTCTAGCTAAAGGATTGTAAATGCAACAAT 1812

RESULT 17
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

  Query Match             12.2%;  Score 161.8;  DB 3;  Length 80246;
  Best Local Similarity   58.7%;  Pred. No. 2.4e-43;
  Matches  360;  Conservative    0;  Mismatches  234;  Indels   19;  Gaps    4;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              ||||| ||  |||   ||||||  || || ||  | | |||  ||  |||    || |||
Db      58640 CAAAACCGCTGAGGCCTAGACTCTCTTACTGCTGAGAAAGGAAGATTCTG-CACTTCTTA 58698

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
              |||| ||| || ||||  |||  ||| || || || || ||    || |      |
Db      58699 GGGGTAGAGTGTTGTTTTTATACTAACCAGTCAGGGATAATATGAGATACCACCCAGTGT 58758

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |    ||       |     |    |     ||||  |||| ||      |    |  |
Db      58759 TTACAGGAAAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCT------TATACCAAC 58812

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||| |    |||  | | |||  |||||||||  |||| ||| |  |||| ||  |  |
Db      58813 CTCTGGAGTTGGGCGACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATCTTGCTA 58872

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
               || ||   ||||| |||||||| || ||| || | || || ||||| || || || ||
Db      58873 ATAGTCGCATTTGGGCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCTAGGATC 58932

Qy        302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
              || ||  | |||||||| ||  |  | |||||| ||||||| |||   || |  |  ||
Db      58933 GAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACTAACAAC 58992

Qy        362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
               ||| | | |||||||||||||| || ||| | |  | |         |||| | |  |
Db      58993 TTCTGCTGAGGACCCCTGGACCGACCCGCTGGCCCTTTC---------AATGGCCTAAAG 59043

Qy        422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
                | ||||||  ||||| |   | ||||||   ||||| || ||| || || ||| ||||
Db      59044 AGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCAGCAGGA 59103

Qy        482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
              || || ||| ||  ||  | ||    |||||||| |  |||| | |||||||  |||
Db      59104 AGTAGCTACAGCGGTCATCGCCAAA---TCCCAACAGCAGCTGGGGTGTCCTGTTTGGAG 59160

Qy        542 GGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTA 601
              || ||| ||||||      | ||||||  |||   |   |     || |     | | |
Db      59161 GGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAGAACTTT 59220

Qy        602 NCTGGGAAGGTGA 614
                ||   || | |
Db      59221 TGTGTCTAGCTAA 59233

RESULT 18
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 80595
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

  Query Match             11.4%;  Score 151.8;  DB 3;  Length 80595;
  Best Local Similarity   58.5%;  Pred. No. 6.5e-40;
  Matches  327;  Conservative    0;  Mismatches  214;  Indels   18;  Gaps    3;

Qy         56 TTTTTAGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTT 115
              || ||||||| ||| || ||||  |||  ||| || || || ||  |    || |
Db      58926 TTCTTAGGGGTAGAGTGTTGTTTTTATACTAACCAGTCAGGGATAGTATGAGATACCACC 58985
```

```
Qy        116 AAAGAAATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGG 175
               |     |    ||       |     |    | |   ||||  |||| ||      |
Db      58986 CAGTGTTTACAGGAAAAGGCTTCTGAAATCAGACAATGCCTTTCAAACTCT------TAT 59039

Qy        176 GGCCTCCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATA 235
                |  |||| |    |||  | | |||  |||||||||  |||| ||| |  |||| ||
Db      59040 ACCAACCTCTGGAGTTGGGCGACATGGCTTCTCCCCTTTCTAGGTCCTGTGACAGCCATC 59099

Qy        236 ATATTTTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCC 295
               |  |  || ||    ||||| |||||||| || ||| || | || || ||||| || ||
Db      59100 TTGCTAATAGTCGCATTTGGGCCCTGTATTTTTAACCTCTTGGTCAAATTTGTTTCCTCT 59159

Qy        296 AGAATTGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATG 355
              || || || ||  | |||||||| ||  |  | |||||| ||||||| |||   || |
Db      59160 AGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAATGAGCTCAACT 59219

Qy        356 ACTAAAATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGAC 415
              |  ||  ||| | | |||||||||||||| || ||| | |          | | |||| |
Db      59220 AACAACTTCTGCTGAGGACCCCTGGACCGACCCGCTGGCC---------CTTTCAATGGC 59270

Qy        416 ATTGAAGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCA 475
               |  |   | ||||||  ||||| |   | ||||||   ||||| || ||| || || ||
Db      59271 CTAAAGAGCTCCCCTCTGGAGGACACTACCACTGCAGGGCCCCTTCTTCACCCCTATCCA 59330

Qy        476 GTAGGAAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGT 535
              | |||||| || || ||| ||  ||  | ||    |||||||| |  |||| | ||||||
Db      59331 GCAGGAAGTAGCTACAGCGGTCATCGCCAAA---TCCCAACAGCAGCTGGGGTGTCCTGT 59387

Qy        536 TGAGAGGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNA 595
              |  ||||| ||| ||||||      | ||||||  |||   | | |     || |
Db      59388 TTGGAGGGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTCAGGTGGGGACTTGGAG 59447

Qy        596 AGCCTANCTGGGAAGGTGA 614
              | | |   ||   || | |
Db      59448 AACTTTTGTGTCTAGCTAA 59466


RESULT 19
US-09-621-976-2599/c
; Sequence 2599, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2599
;   LENGTH: 506
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 67..228
US-09-621-976-2599

  Query Match              9.3%;  Score 123.6;  DB 4;  Length 506;
  Best Local Similarity   82.4%;  Pred. No. 8.7e-32;
  Matches  140;  Conservative   10;  Mismatches   19;  Indels    1;  Gaps    1;

Qy        923 CATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTTCCATCCCTTTGGAT 982
              | ||||||||||||||||||||||  | |||||| || ||||||| |||||||| |||||
Db        169 CGTCCACCACTGCTGTTTGCCACCACCGCAGACCTGCCGCTGACTCCCATCCCTCTGGAT 110

Qy        983 CCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCATTGCCACTCCCGATCAG 1042
              || |||| |||| | ||||||||||||||||||||||  ||||||||| ||||| ||  |
Db        109 CCTGCAGGGTGT-CGCTGTGCTCCTGATCCAGCGAGGCGCCCATTGCCGCTCCCAATTGG 51

Qy       1043 GCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTCCT 1092
              |||||||||||||||||||||||||| ||||||||:::::|::|:|: :|
Db         50 GCTAAAGGCTTGCCATTGTTCCTGCACGGCTAAGTKMMKKGSYTKGKTYT 1


RESULT 20
US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
;  APPLICANT: YOON, JI-WON
;  APPLICANT:  JUN, HEE-SOOK
;  APPLICANT:  PARK, HAE-JOON
;  APPLICANT:  AHN, JONG-SEONG
;  APPLICANT:  HA, YOUNG-JU
;  APPLICANT:  CHUNG, SOO-IL
;  TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
;  FILE REFERENCE: 98-338
;  CURRENT APPLICATION NUMBER: US/09/120,653D
;  CURRENT FILING DATE:  1998-07-22
;  PRIOR APPLICATION NUMBER: KR 98-10108
;  PRIOR FILING DATE: 1998-03-24
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 1
;   LENGTH: 3910
;   TYPE: DNA
;   ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

  Query Match              8.1%;  Score 107.8;  DB 3;  Length 3910;
  Best Local Similarity   56.5%;  Pred. No. 9.6e-26;
  Matches  308;  Conservative    0;  Mismatches  217;  Indels   20;  Gaps    5;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              ||||| ||  |||   ||||  | || || ||  ||  | | |||  |||    || |||
Db       2948 CAAAACCGCTGAGGCATAGATGTCCTCACTGCTGAAGAAAGAGGACTCTGCACTTTCTTA 3007

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
              |||||||| || ||||   |   ||| ||  |  | ||      |||||  |
Db       3008 GGGGAAGAGTGTTGTTTTGACACTAACCAGGCACGGATAGCA--TGAGATGGCACCCAGC 3065

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |||       ||        | |  ||  | | || |||    |  ||       |  |
Db       3066 GTTTACAGAAAAAGGCTGCTGAAATGAGTCGCCTTTTAAATTCTTATAC-------CAAC 3118

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||  |     ||    | |||  |||| ||||  |||| ||  | ||| |  |  |  |
Db       3119 CTGTGGTGTGGGGCAACATGGCTTCTCACCTTTCTAGGTCCCGTGGCAACCGTCTTGGTG 3178

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
              ||||||  |||||| ||| | || || ||| | ||||| || ||||| |  || ||| |
Db       3179 TTACTCGCCTTTGGGCCCAGCATTTTTAACGTTCTTGTCAAATTTGT-TTGTCTAGATTC 3237

Qy        302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
              || ||  | |||||||| || ||  | || ||  ||||||  ||    || |  |   |
Db       3238 GAGGCCTTCAAGCTACAGATGGTCTTACATATCAAACCCCGAATAATTTCAACTAACTAC 3297

Qy        362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
               ||||| | |||| |||||||   || ||| |    | | |||          | | ||
Db       3298 TTCTACGGAGGACACCTGGACTAACCAGCTGGCAGTTACCCTGG---------CCTAGAG 3348

Qy        422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
                  ||||||   |||  |   |||||||| | |||||  | | | || || ||  ||||
Db       3349 AGTTCCCCTCTGAAGGTCACTACAACTGCAAAGCCCCTTTTTCGCGCCTATCCATCAGGA 3408

Qy        482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
               | |  |||| ||||  || ||| |  |  ||||||| | ||||  | |||||||||  |
Db       3409 CGTACCTAGAACAGTCCTCGGCC-ATATTGCCAACAGCAGTTGGAGTGTCCTGTTGATTG 3467

Qy        542 GGTGG 546
               | ||
Db       3468 AGGGG 3472
```

```
RESULT 21
US-09-513-999C-24228
; Sequence 24228, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 24228
;   LENGTH: 149
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 127
;   OTHER INFORMATION: r=a or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 129
;   OTHER INFORMATION: n=a, g, c or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 135
;   OTHER INFORMATION: m=a or c
US-09-513-999C-24228

  Query Match             7.8%;  Score 103.2;  DB 4;  Length 149;
  Best Local Similarity   87.1%;  Pred. No. 3.7e-25;
  Matches  122;  Conservative    2;  Mismatches   15;  Indels    1;  Gaps    1;

Qy        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
              | ||||||||| ||| ||||||||||||||||||||||||||||  ||||| |||| |||
Db          5 CTCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTATAAAACTA 64

Qy        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
              ||||| ||||| |||||||| ||||||||||||||||||||||| |||||| | |||||
Db         65 CAAATGGTTCTCCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCTACTGCGGACCT 124

Qy        377 CTGG-ACCGGCCTGCTAGAC 395
              ||:| |||||:||||||| |
Db        125 CTRGNACCGGMCTGCTAGCC 144

RESULT 22
US-09-799-451-249
; Sequence 249, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Goodrich, Ryle
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Yamazaki, Victoria
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Ghosh, Reena
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6783969el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 803
;  CURRENT APPLICATION NUMBER: US/09/799,451
;  CURRENT FILING DATE:  2001-03-05
;  NUMBER OF SEQ ID NOS: 948
;  SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 249
;   LENGTH: 2908
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (382)..(1995)
US-09-799-451-249

  Query Match             6.5%;  Score 85.8;  DB 4;  Length 2908;
  Best Local Similarity   55.1%;  Pred. No. 2.7e-18;
  Matches  190;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||  |||   ||||| || |||| ||   |   || ||||  |||||    ||
Db       1623 TCAAAATCGTCGAGGACTAGACATGTTAACGGCAGCACAGGGAGGAATTTGTTTGGCCTT 1682

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||  ||| |||| || |  |  || |||||||| ||||   |    || ||  | | | |
Db       1683 AGATGAAAAATGTTGCTTTTGGGTAAATCAATCAGGAAAAGTACAAGACAACATCAGACA 1742

Qy        121 AATTTGAGATCGAATATAATGT---AGAGCAGAGGACCTTCAAAACACTGCACCCTGGGG 177
              | |   | ||| |   |   ||    | | |   | |  |||         |   ||||
Db       1743 ACTCCTAAATCAAGCCTCCAGTTTACGGGAACGAGCCACTCAGGGTTGGTTAAATTGGGA 1802

Qy        178 CCTCCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAAT 237
                      |  ||||| |  | |||  ||| ||| |   ||| || || |    | |  |
Db       1803 AGGAACTTGGAAATGGTTCTCTTGGGTTCTTCCCCTTACAGGCCCACTTGTTAGTCTCCT 1862

Qy        238 ATTTTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAG 297
              | |||| ||||| ||||| || ||| || | ||  |  |     | |||||||| ||  |
Db       1863 ACTTTTGCTCCTTTTTGGTCCATGTCTCCTAAATCTAATAACCCAATTTGTCTCCTCTCG 1922

Qy        298 AATTGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCA 342
                || | ||  ||||||| || |    |||       | || |||
Db       1923 CCTTCAGGCCATAAAGCTCCAGACGAATCTCAGTGCAGGACGCCA 1967

RESULT 23
US-08-007-282B-1
; Sequence 1, Application US/08007282B
; Patent No. 5403582
;  GENERAL INFORMATION:
;    APPLICANT:  NAZERIAN, KEYVAN
;    APPLICANT:  CALVERT, JAY G.
;    APPLICANT:  WITTER, RICHARD L.
;    APPLICANT:  YANAGIDA, NOBORU
;    TITLE OF INVENTION:  VACCINE COMPRISING FOWLPOX VIRUS
;    TITLE OF INVENTION:  RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF AN
;    TITLE OF INVENTION:  AVIAN RETICULOENDOTHELIOSIS RETROVIRUS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  8110 Gatehouse Road  Suite 500 East
;      CITY:  Falls Church
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/007,282B
;      FILING DATE:  19930121
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY JR., GERALD M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1644-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1704 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-007-282B-1

  Query Match              6.1%;  Score 80.6;  DB 1;  Length 1704;
  Best Local Similarity   53.7%;  Pred. No. 1.1e-16;
  Matches  167;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              |||||| ||||||  |||||| |  | || ||| ||  ||| ||||  ||| |     |
Db       1387 CAAAATAGAAGAGGGTTAGACCTATTGACTGCCGAACAAGGAGGAATATGTCTCGCACTC 1446

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
                |||  | || ||||  || | |||  | || || ||| |   ||| ||  |   | ||
Db       1447 CAGGAGAAGTGTTGTTTTTACGCTAACAAGTCGGGTATCGTACGTGACAAGATCCGAAAA 1506

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |   |||     |       |||  |   |  ||  |  |||   | |  ||| ||  |
Db       1507 CTCCAAGAGGACCTTATCGAGAGAAAACGTGCACTGTACGACAACCCCCTGTGGAGCGGC 1566

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
               | | |   |   | || |     || ||||| ||||| || ||    |   | |||||
Db       1567 TTGAACGGCTTCCTTCCATATTTGCTACCCTTGTTAGGCCCCCTGTTTGGGCTCATATTG 1626

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
              || ||   | | || || || ||  | ||   |||   |    || |   |  || ||||
Db       1627 TTCCTGACCCTCGGCCCGTGCATTATGAAGACCCTGACTCGCATTATACATGACAAAATT 1686

Qy        302 GAAGCTGTAAA 312
               | || |||||
Db       1687 CAGGCAGTAAA 1697

RESULT 24
US-09-011-745-1
; Sequence 1, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2518
```

```
;   TYPE: DNA
;   ORGANISM: RD114
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-1

  Query Match              6.1%;  Score 80.6;  DB 3;  Length 2518;
  Best Local Similarity   53.7%;  Pred. No. 1.5e-16;
  Matches  167;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              |||||| | || |   | ||| | ||||| ||  ||  ||| ||||  ||||||   |||
Db       1930 CAAAATAGGAGGGGACTGGACCTACTAACGGCAGAACAAGGAGGAATTTGTTTAGCCTTA 1989

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
                 ||| |||||||||  |||| |||  | || |||||  | |   | ||| | | |
Db       1990 CAAGAAAAATGCTGTTTTTATGCTAACAAGTCAGGAATTGTGAGAAACAAAATAAGAACC 2049

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |   |||   | || ||    |    ||   |||   || ||   | | ||||  |
Db       2050 CTACAAGAAGAATTACAAAAACGCAGGGAAAGCCTGGCAACCAACCCTCTCTGGACCGGG 2109

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||       |   | || |     || ||  || | ||||| |||    |  |  || |
Db       2110 CTGCAGGGCTTTCTTCCGTACCTCCTACCTCTCCTGGGACCCCTACTCACCCTCCTACTC 2169

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
               ||||   | |||| || ||  | ||||    |||  |    ||  |   |   ||| ||
Db       2170 ATACTAACCATTGGGCCATGCGTTTTCAGTCGCCTCATGGCCTTCATTAATGATAGACTT 2229

Qy        302 GAAGCTGTAAA 312
               | | |||| |
Db       2230 AATGTTGTACA 2240

RESULT 25
US-09-555-352-1
; Sequence 1, Application US/09555352
; Patent No. 6544779
; GENERAL INFORMATION:
;  APPLICANT: Cichutek, Klaus
;  APPLICANT:  Merget-Millitzer, Heike
;  TITLE OF INVENTION: PSEUDO-TYPE RETROVIRAL VECTORS WITH
;  TITLE OF INVENTION:  MODIFIABLE SURFACE CAPSID PROTEINS
;  FILE REFERENCE: 11692-005001
;  CURRENT APPLICATION NUMBER: US/09/555,352
;  CURRENT FILING DATE:  2000-08-24
;  PRIOR APPLICATION NUMBER: PCT/DE98/03542
;  PRIOR FILING DATE: 1998-11-27
;  PRIOR APPLICATION NUMBER: DE 197 52 855.4
;  PRIOR FILING DATE: 1997-11-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 4776
;   TYPE: DNA
;   ORGANISM: Murine leukemia virus
US-09-555-352-1

  Query Match              6.1%;  Score 80.6;  DB 4;  Length 4776;
  Best Local Similarity   53.7%;  Pred. No. 2.4e-16;
  Matches  167;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              |||||| ||||||  |||||| |  | || ||| ||  ||| ||||  ||| |     |
Db       1406 CAAAATAGAAGAGGGTTAGACCTATTGACTGCCGAACAAGGAGGAATATGTCTCGCACTC 1465

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
                |||  | || ||||  || | |||  | || || ||| |   ||| ||  |   | ||
Db       1466 CAGGAGAAGTGTTGTTTTTACGCTAACAAGTCGGGTATCGTACGTGACAAGATCCGAAAA 1525
```

```
Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |   |||     |       |||  |   |  ||  |  |||   | |  ||| ||  |
Db       1526 CTCCAAGAGGACCTTATCGAGAGAAAACGTGCACTGTACGACAACCCCCTGTGGAGCGGC 1585

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
               | | |   |   | || |     || ||||| ||||| || ||    |   | |||||
Db       1586 TTGAACGGCTTCCTTCCATATTTGCTACCCTTGTTAGGCCCCCTGTTTGGGCTCATATTG 1645

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
              || ||   | | || || || ||  | ||   |||   |    || |   |  || ||||
Db       1646 TTCCTGACCCTCGGCCCGTGCATTATGAAGACCCTGACTCGCATTATACATGACAAAATT 1705

Qy        302 GAAGCTGTAAA 312
               | || |||||
Db       1706 CAGGCAGTAAA 1716

RESULT 26
US-09-011-745-8
; Sequence 8, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 5865
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3611)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3612)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3613)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3614)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3799)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3800)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3801)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (3802)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-8

  Query Match             6.1%;  Score 80.6;  DB 3;  Length 5865;
  Best Local Similarity   53.7%;  Pred. No. 2.7e-16;
  Matches  167;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              |||||| | || |   | ||| | ||||| ||  ||  ||| ||||  ||||||   |||
Db       2788 CAAAATAGGAGGGGACTGGACCTACTAACGGCAGAACAAGGAGGAATTTGTTTAGCCTTA 2847

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
                 ||| |||||||||  |||| |||  | || |||||  | |   | ||| | | |
Db       2848 CAAGAAAAATGCTGTTTTTATGCTAACAAGTCAGGAATTGTGAGAAACAAAATAAGAACC 2907

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |   |||   | || ||    |    ||   |||   || ||   | | ||||  |
Db       2908 CTACAAGAAGAATTACAAAAACGCAGGGAAAGCCTGGCAACCAACCCTCTCTGGACCGGG 2967

Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||       |   | || |     || ||  || | ||||| |||    |  |  || |
Db       2968 CTGCAGGGCTTTCTTCCGTACCTCCTACCTCTCCTGGGACCCCTACTCACCCTCCTACTC 3027

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
               ||||   | |||| || ||  | ||||    |||  |    ||  |   |   ||| ||
Db       3028 ATACTAACCATTGGGCCATGCGTTTTCAGTCGCCTCATGGCCTTCATTAATGATAGACTT 3087

Qy        302 GAAGCTGTAAA 312
               | | |||| |
Db       3088 AATGTTGTACA 3098

RESULT 27
US-09-120-653D-6
; Sequence 6, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
;  APPLICANT: YOON, JI-WON
;  APPLICANT:  JUN, HEE-SOOK
;  APPLICANT:  PARK, HAE-JOON
;  APPLICANT:  AHN, JONG-SEONG
;  APPLICANT:  HA, YOUNG-JU
;  APPLICANT:  CHUNG, SOO-IL
;  TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
;  FILE REFERENCE: 98-338
;  CURRENT APPLICATION NUMBER: US/09/120,653D
;  CURRENT FILING DATE:  1998-07-22
;  PRIOR APPLICATION NUMBER: KR 98-10108
;  PRIOR FILING DATE: 1998-03-24
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 6
;   LENGTH: 464
;   TYPE: DNA
;   ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-6

  Query Match             5.9%;  Score 78.2;  DB 3;  Length 464;
  Best Local Similarity   55.9%;  Pred. No. 3.1e-16;
  Matches  213;  Conservative    0;  Mismatches  158;  Indels   10;  Gaps    3;

Qy          2 CAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTTA 61
              ||||| ||  |||   ||||  | || || ||  ||  | | |||  |||    || |||
Db         91 CAAAACCGCTGAGGCATAGATGTCCTCACTGCTGAAGAAAGAGGACTCTGCACTTTCTTA 150

Qy         62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
              |||||||| || ||||   |   ||| ||  |  | ||      |||||  |
Db        151 GGGGAAGAGTGTTGTTTTGACACTAACCAGGCACGGATAGCA--TGAGATGGCACCCAGC 208

Qy        122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTC 181
               |||       ||        | |  ||  | | || |||    |  ||       |  |
Db        209 GTTTACAGAAAAAGGCTGCTGAAATGAGTCGCCTTTTAAATTCTTATAC-------CAAC 261
```

```
Qy        182 CTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
              ||  |     ||    | |||  |||| ||||  |||| ||  | ||| |  |  |  | 
Db        262 CTGTGGTGTGGGGCAACATGGCTTCTCACCTTTCTAGGTCCCGTGGCAACCGTCTTGGTG 321

Qy        242 TTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATT 301
              ||||||  |||||| ||| | || || ||| | ||||| || ||||| |  || ||| | 
Db        322 TTACTCGCCTTTGGGCCCAGCATTTTTAACGTTCTTGTCAAATTTGT-TTGTCTAGATTC 380

Qy        302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
              || ||  | ||||||||| || ||  | || ||  ||||||  ||    || |  |   | 
Db        381 GAGGCCTTCAAGCTACAGATGGTCTTACATATCAAACCCCGAATAATTTCAACTAACTAC 440

Qy        362 ATCTACCGTGGACCCCTGGAC 382
               ||||| | |||| |||||||
Db        441 TTCTACGGAGGACACCTGGAC 461

RESULT 28
US-09-621-976-11057
; Sequence 11057, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 11057
;   LENGTH: 492
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-11057

  Query Match              5.7%;  Score 76;  DB 4;  Length 492;
  Best Local Similarity   59.2%;  Pred. No. 1.8e-15;
  Matches  190;  Conservative    0;  Mismatches  116;  Indels   15;  Gaps    3;

Qy        280 TAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACC 339
              |||    || |  || || ||  | ||  | ||||| ||||| |   | ||||||||||||
Db         95 TAATCAAGTGTAATCAAGGATCAAGGCCATCAAGCTGCAAATGGACTTACAAATGGAACC 154

Qy        340 CCAGATGCAGTCCATGACTAAAATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATG 399
               || |||   || |         ||||| | | ||||||||| |  |||||| | |  | 
Db        155 TCAAATGAGCTCAACTCACGGCTTCTACTGAGAACCCCTGGATCCACCTGCTGGTCCCTC 214

Qy        400 CTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCT 459
                        | |  | |  ||    ||||||  ||||| | | | ||||||||   |||||
Db        215 ----------ACTAGCCTAAAAAGTTCCCCTCTGGAGGACACCACAACTGCAGGGCCCCT 264

Qy        460 ACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGT 519
               || ||| || |  ||  |||||| ||  ||| |   ||    |     | |||||||| 
Db        265 TCTTCACCCCTAACCATCAGGAAGTAGCCAGAACGACTGCTGCC---AGTTCCCAACAGC 321

Qy        520 ACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGC 579
              | ||||  | ||||||  |||| |||||  |||   |  | | ||||||  ||||| || 
Db        322 AATTGGTGTGTCCTGTCTAGAGGGGGGACT--GAGGAGGTTCCAGCTGGGCTTCCTGGGT 379

Qy        580 TGACTAAGAATCCCNAAGCCT 600
               || || |    |  ||  ||
Db        380 CGAGTAGGGGCTCAGAAAGCT 400

RESULT 29
US-09-621-976-17982
; Sequence 17982, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 17982
;   LENGTH: 442
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-17982

  Query Match              5.5%;  Score 73.4;  DB 4;  Length 442;
  Best Local Similarity   83.0%;  Pred. No. 1.3e-14;
  Matches   83;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       1230 AANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGCCATTTTGGTAGCGGCCCACCA 1289
              ||    || ||||||||||| | |||||  ||||||  |||| |||| || |||||||||
Db          1 AACACAAGCCTTGCCACCATCTTGGAAGCAGCCCACCACCATCTTGGAAGTGGCCCACCA 60

Qy       1290 CCATCTTGGGAGCTGTGGGAGCAAGGATCCCCCAGTAACA 1329
              |||||||||||||| |||||||||||| |||||||||| |
Db         61 CCATCTTGGGAGCTCTGGGAGCAAGGACCCCCCAGTAAAA 100

RESULT 30
US-09-620-312D-171
; Sequence 171, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  John Tillinghast
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6569662el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/09/620,312D
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1105
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 171
;   LENGTH: 2809
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1936)..(2364)
US-09-620-312D-171

  Query Match              5.2%;  Score 69.6;  DB 4;  Length 2809;
  Best Local Similarity   55.8%;  Pred. No. 9.6e-13;
  Matches  154;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;
```

```
Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||  | ||||   | ||| | |||||| |  | | |||||| |||||  |||   |
Db       1995 TCAAAACTGGAGAGTCCTGGACCTCCTAACCACTGAGAAAGGGGGTACCTGCATATACCT 2054

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
                 ||||||||||||||  | ||||||| ||||||| ||  ||  |     ||||
Db       2055 CCAGGAAGAATGCTGTTTCTGTGTTAATGAATCTGGCATTGTTCATATCGCAGTTCGTAG 2114

Qy        121 AATTTGAGATCG---AATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGG 177
                ||   ||  |      | |   ||||    | | |  | |   |      |   |
Db       2115 GCTTCATGACAGGGCTGCAGAGCTTAGACATCAAGTCGCTGACTCCTGGTGGCAAGGATC 2174

Qy        178 CCTCCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAAT 237
                 |||      |||||| ||||||  |  ||||||| ||||||| ||        |  |
Db       2175 ATCCCTTCTAAGATGGATACCCTGGGTTGCCCCCTTCCTAGGACCCCTGATCTTCCTCTT 2234

Qy        238 ATTTTTACTCCTCTTTGGACCCTGTATCTTCAACTT 273
                | |||||  |  |||| || || || || ||| |
Db       2235 CCTGTTACTAATGATTGGGCCATGCATATTTAACCT 2270

Search completed: January 22, 2005, 19:37:03
Job time : 127.063 secs
```

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          January 22, 2005, 12:12:40 ; Search time 721.788 Seconds
                                              (without alignments)
                                              10579.682 Million cell updates/sec

Title:           US-09-319-156B-12
Perfect score:   1329
Sequence:        1 tcaaaatcgaagagctttag..........gcaaggatcccccagtaaca 1329

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters:      8600550

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                 1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                 2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                 3:  /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                 4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                 5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                 6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                 7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                 8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                 9:  /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                 10:  /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
                 11:  /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
                 12:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                 13:  /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
                 14:  /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
                 15:  /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
                 16:  /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
                 17:  /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
                 18:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                 19:  /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
                 20:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                 21:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                    SUMMARIES
                   %
Result           Query
  No.   Score    Match Length DB  ID                       Description
-----------------------------------------------------------------------------
    1    1326    99.8   1329  8  US-08-979-847-108          Sequence 108, App
    2    1326    99.8   1329  15  US-10-114-104-108          Sequence 108, App
    3  1035.2    77.9  56093  10  US-09-873-367C-81          Sequence 81, Appl
    4   798.2    60.1  21646  13  US-10-087-192-910          Sequence 910, App
    5   778.8    58.6   1393  16  US-10-220-120-17           Sequence 17, Appl
    6   775.4    58.3   2030  16  US-10-637-565-18           Sequence 18, Appl
    7   774.2    58.3   7974  14  US-10-198-846-9936         Sequence 9936, Ap
    8   766.8    57.7   2074  16  US-10-416-642-4            Sequence 4, Appli
    9   731.8    55.1   2946  13  US-10-114-893-134          Sequence 134, App
   10   731.8    55.1   2946  15  US-10-016-249-3            Sequence 3, Appli
   11   730.8    55.0   2930  9  US-09-902-535-1            Sequence 1, Appli
   12   722.6    54.4   7582  16  US-10-632-793-30           Sequence 30, Appl
   13   721.2    54.3   1136  16  US-10-632-793-25           Sequence 25, Appl
   14   712.6    53.6   2782  16  US-10-632-793-26           Sequence 26, Appl
   15     711    53.5   2782  16  US-10-133-036-1            Sequence 1, Appli
   16   704.8    53.0   1894  9  US-09-864-761-4444        Sequence 4444, Ap
c  17   643.4    48.4 161334  13  US-10-087-192-730          Sequence 730, App
   18   637.8    48.0    780  10  US-09-854-867-385          Sequence 385, App
c  19   601.4    45.3 157090  17  US-10-672-764A-34          Sequence 34, Appl
c  20   579.8    43.6 180557  13  US-10-003-806-6            Sequence 6, Appli
c  21   579.8    43.6 180557  13  US-10-003-806-9            Sequence 9, Appli
   22   570.2    42.9 161671  15  US-10-017-117-1            Sequence 1, Appli
   23   545.4    41.0 134292  16  US-10-240-425-1102         Sequence 1102, Ap
   24     545    41.0    635  8  US-08-979-847-102         Sequence 102, App
   25     545    41.0    635  15  US-10-114-104-102          Sequence 102, App
   26   536.2    40.3    849  16  US-10-220-120-15           Sequence 15, Appl
c  27   530.6    39.9 366710  18  US-10-719-993-6792         Sequence 6792, Ap
c  28     512    38.5  17758  16  US-10-264-237-2833         Sequence 2833, Ap
c  29     502    37.8 326014  9  US-09-731-231A-3          Sequence 3, Appli
c  30     502    37.8 326014  17  US-10-751-985-3            Sequence 3, Appli
   31   478.6    36.0 283351  18  US-10-719-993-7065         Sequence 7065, Ap
   32   478.6    36.0 1980090  18  US-10-719-993-6815         Sequence 6815, Ap
   33   468.8    35.3   2052  16  US-10-276-774-678          Sequence 678, App
   34     465    35.0   3372  16  US-10-632-793-28           Sequence 28, Appl
   35     465    35.0  22436  11  US-09-997-722-148          Sequence 148, App
   36     460    34.6 285020  13  US-10-087-192-1666         Sequence 1666, Ap
c  37   457.8    34.4   2145  13  US-10-027-632-98772        Sequence 98772, A
c  38   457.8    34.4   2145  15  US-10-027-632-98772        Sequence 98772, A
   39   455.8    34.3   1003  16  US-10-637-565-1            Sequence 1, Appli
   40   455.8    34.3   2052  16  US-10-637-565-14           Sequence 14, Appl
c  41   453.6    34.1    559  9  US-09-864-761-7501        Sequence 7501, Ap
   42     444    33.4    581  15  US-10-029-386-5948         Sequence 5948, Ap
   43   440.4    33.1   8523  10  US-09-854-867-21           Sequence 21, Appl
   44   440.2    33.1   1321  16  US-10-632-793-20           Sequence 20, Appl
   45     438    33.0   2349  15  US-10-104-047-1148         Sequence 1148, Ap

                                    ALIGNMENTS


RESULT 1
US-08-979-847-108
; Sequence 108, Application US/08979847
; Publication No. US20030039664A1
;  GENERAL INFORMATION:
;    APPLICANT:  PERRON, HERVE
;    APPLICANT:  BESEME, FREDERIC
;    APPLICANT:  BEDIN, FREDERIC
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  KOMURIAN-PRADEL, FLORENCE
;    APPLICANT:  JOLIVET-REYNAUD, COLETTE
;    APPLICANT:  MANDRAND, BERNARD
;    APPLICANT:  GARSON, JEREMY
;    APPLICANT:  TUKE, PHILIP
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  210
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/979,847
;      FILING DATE:  26-NOV-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  BERRIDGE, WILLIAM P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 39046A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  108:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1329 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-979-847-108

  Query Match             99.8%;  Score 1326;  DB 8;  Length 1329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1329;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260

Qy       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

Qy       1321 CCAGTAACA 1329
              |||||||||
Db       1321 CCAGTAACA 1329


RESULT 2
US-10-114-104-108
; Sequence 108, Application US/10114104
; Publication No. US20030198647A1
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/114,104
;            FILING DATE: 03-Apr-2002
;            CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/979,847
;              FILING DATE: 26-NOV-1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: BERRIDGE, WILLIAM P.
;              REGISTRATION NUMBER: 30,024
;              REFERENCE/DOCKET NUMBER: WPB 39046A
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 703-836-6400
;              TELEFAX: 703-836-2787
;    INFORMATION FOR SEQ ID NO: 108:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1329 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 108:
US-10-114-104-108

  Query Match             99.8%;  Score 1326;  DB 15;  Length 1329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1329;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900

Qy         901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960

Qy         961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

Qy        1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080

Qy        1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140

Qy        1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200

Qy        1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGG 1260

Qy        1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 CCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

Qy        1321 CCAGTAACA 1329
               |||||||||
Db        1321 CCAGTAACA 1329


RESULT 3
US-09-873-367C-81
; Sequence 81, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  APPLICANT:  Soppet, Daniel
;  APPLICANT:  Endress, Gregory
;  APPLICANT:  Augustus, Meena
;  APPLICANT:  Ebner, Reinhard
;  APPLICANT:  Carter, Kenneth
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
;  TITLE OF INVENTION:  Signature Gene Sets
;  FILE REFERENCE: 689290-64
;  CURRENT APPLICATION NUMBER: US/09/873,367C
;  CURRENT FILING DATE:  2003-04-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,891
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,842
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/244,867
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: U.S. 60/245,084
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 1067
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 81
;   LENGTH: 56093
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-873-367C-81

  Query Match             77.9%;  Score 1035.2;  DB 10;  Length 56093;
  Best Local Similarity   90.1%;  Pred. No. 9.7e-311;
  Matches 1133;  Conservative    0;  Mismatches  111;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db      37018 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 37077

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db      37078 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 37137

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db      37138 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 37197

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      37198 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 37257

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db      37258 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 37317

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db      37318 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 37365

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db      37366 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 37425

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db      37426 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 37485

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db      37486 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 37545

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      37546 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 37605

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db      37606 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 37665

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db      37666 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 37725

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db      37726 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 37785

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db      37786 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 37845

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGG 900
              ||  |||||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||
Db      37846 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 37905

Qy        901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCT 960
              ||  ||||||||||| | ||||||||||||||||||||||||| ||  | |||||||||
Db      37906 CTTGAGCTGAGCTTTCGCTCGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGCC 37965

Qy        961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
              ||||||| |||||||| ||||||  |||| |||||| ||||||||||||||||||||||
Db      37966 GCTGACTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGC 38025

Qy       1021 ACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 1080
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||||||||
Db      38026 ACCCATTGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCC 38085

Qy       1081 TGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGAC 1140
              ||||||  ||||||| || ||||||||| |||||||||||||||||||||||||  ||||
Db      38086 TGGGTTCATCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 38145

Qy       1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
              |||| ||||||||||||||||||||||||||||||||||||||| ||||||||||| | |
Db      38146 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 38204

Qy       1201 TCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              ||  | |||||||||||||||||||||||||   ||||||||||||||| | || ||
Db      38205 TCCATAAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 38261


RESULT 4
US-10-087-192-910
; Sequence 910, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 910
;   LENGTH: 21646
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(21646)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-910

  Query Match             60.1%;  Score 798.2;  DB 13;  Length 21646;
  Best Local Similarity   87.1%;  Pred. No. 5.1e-237;
  Matches  908;  Conservative    0;  Mismatches  106;  Indels   28;  Gaps    2;

Qy        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
              ||||||||||| ||| |||||||||||||||||||||||||||||||| ||||||| |||
Db      19571 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAACTGTAAAACTA 19630

Qy        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
              ||||| |||||||||||||| ||||||||||||||||||||||| ||| |||   |||||
Db      19631 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACCACAGACCC 19690

Qy        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436
              ||||||| ||||||||| | |||||| |||||||||||||| |||| |||||| ||||||
Db      19691 CTGGACCAGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCCCCCGAG 19750

Qy        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496
              |||||||||||||  ||||||||||||  | ||||||||| |||||||||||| ||| ||
Db      19751 GAAATCTCAACTGTGCAACCCCTACTATGCCCCAATTCAGCAGGAAGCAGTTACAGCGGT 19810

Qy        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556
                |||||||||||||||||||| |||||||||||||||||||||||| | ||||||||||
Db      19811 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGCACTGAGAGAC 19870

Qy        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616
              | ||||||||||||||||||||| ||||||||||||| |||| || |||||||||||||
Db      19871 AAGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCGTAGCTGGGAAGGTGACT 19930

Qy        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676
              ||||||| ||||||||||||||||||| ||||||||||||||| ||||||||||||||||
Db      19931 GCATCCACCTTTAAACATGGGGCTTGCCACTTAGCTCACACCCAACCAATCAGAGAGCTC 19990

Qy        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736
              |||||||||||||| ||||||||||||||||||||| |||||||||||| || |||||||
Db      19991 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCACCTTTTGCCTG 20050

Qy        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796
              |||||||||| ||| |||||||||| ||||||||||| | |||||||||||| ||||  |
Db      20051 AGAGCACAGCAGGAGGGACAAGGATCGGGATATAAACCCGGGCATTCAAGCCGGCAATGG 20110

Qy        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856
              ||||||||||||||||||||||| |||||||||||||||||          |||||||||
Db      20111 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT----------TTTCACTCT 20160

Qy        857 ATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCTGAGCTTTT 916
              |||||||| |||||||||||||||||||||| ||||| ||| ||||| |||||||||||
Db      20161 ATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTGTTACGGCTCGAGCTGAGCTTTC 20220

Qy        917 GTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGA------------------CCCG 958
              | || || ||||||||||||||||||||| ||| ||||                  || |
Db      20221 GCTCCCCGTCCACCACTGCTGTTTGCCACTGTCGCAGACCTGCCGCTGACTCCCATCCTG 20280

Qy        959 CTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAG 1018
              ||||||||| ||||||||   ||||| |||| |||||  |||||||||||| |||| |||
Db      20281 CTGCTGACTCCCATCCCTCCAGATCCGGCAGGGTGTCTGCTGTGCTCCTGACCCAGTGAG 20340

Qy       1019 GTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTG 1078
                 |||||||||||||||||||  ||||||||||| ||||||||| |||| |||||||||
Db      20341 ACTCCCATTGCCACTCCCGATCGTGCTAAAGGCTTCCCATTGTTCTTGCACGGCTAAGTG 20400

Qy       1079 CCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATG 1138
              |||||||| |||||||| || ||||||||| ||||||||||||||  |||||||||| ||
Db      20401 CCTGGGTTCGTCCTAATCGAGCTGAACACTAGTCACTGGGTTCCACAGTTCTCTTCCGTG 20460

Qy       1139 ACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGG 1198
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      20461 ACCCACGGCTTCTAATAGAGCTATAACACTCATCGCATGGCCCAAGATTCCATTCCTTGG 20520

Qy       1199 TATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGT 1258
               ||| |||||||||||||||||| |||||||||    |||||||||||||| | |||||
Db      20521 AATCCGTGAGGCCAAGAACCCCAAGTCAGAGAACACAAGGCTTGCCACCATCTTGGAAGC 20580

Qy       1259 GGCCCACTGCCATTTTGGTAGC 1280
              ||| |||  |||| |||| |||
Db      20581 GGCTCACCACCATCTTGGGAGC 20602

RESULT 5
US-10-220-120-17
; Sequence 17, Application US/10220120
; Publication No. US20040048253A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  PANZER, Scott R.
;  APPLICANT:  SPIRO, Peter A.
;  APPLICANT:  BANVILLE, Steven C.
;  APPLICANT:  SHAH, Purvi
;  APPLICANT:  CHALUP, Michael S.
;  APPLICANT:  CHANG, Simon C.
;  APPLICANT:  CHEN, Alice
;  APPLICANT:  D'SA, Steven A.
;  APPLICANT:  AMSHEY, Stefan
;  APPLICANT:  DAHL, Christopher R.
;  APPLICANT:  DAM, Tam C.
;  APPLICANT:  DANIELS, Susan E.
;  APPLICANT:  DUFOUR, Gerard E.
;  APPLICANT:  FLORES, Vincent
;  APPLICANT:  FONG, Willy T.
;  APPLICANT:  GREENAWALT, Lila B.
;  APPLICANT:  HILLMAN, Jennifer L.
;  APPLICANT:  JONES, Anissa L.
;  APPLICANT:  LIU, Tommy F.
;  APPLICANT:  ROSEBERRY, Ann M.
;  APPLICANT:  ROSEN, Bruce H.
;  APPLICANT:  RUSSO, Frank D.
;  APPLICANT:  STOCKDREHER, Theresa K.
;  APPLICANT:  DAFFO, Abel
;  APPLICANT:  WRIGHT, Rachel J.
;  APPLICANT:  YAP, Pierre E.
;  APPLICANT:  YU, Jimmy Y.
;  APPLICANT:  BRADLEY, Diana L.
;  APPLICANT:  BRATCHER, Shawn R.
;  APPLICANT:  CHEN, Wensheng
;  APPLICANT:  COHEN, Howard J.
;  APPLICANT:  HODGSON, David M.
;  APPLICANT:  LINCOLN, Stephen E.
;  APPLICANT:  JACKSON, Stuart
;  TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PT-1113 PCT
;  CURRENT APPLICATION NUMBER: US/10/220,120
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774;
;    60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;
;    60/184,769; 60/184,768; 60/184,837; 60/184,697; 60/184,841;
;    60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;
;    60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;
;    60/204,525; 60/205,285; 60/205,232; 60/205,323; 60/205,287;
;    60/205,324; 60/205,286
;  PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-05-17;
;    2000-05-17; 2000-05-12; 2000-05-16; 2000-05-16; 2000-05-15;
;    2000-05-16; 2000-05-17; 2000-05-16; 2000-05-17; 2000-05-17;
;    2000-05-17; 2000-05-17
;  NUMBER OF SEQ ID NOS: 422
;  SOFTWARE: PERL Program
; SEQ ID NO 17
;   LENGTH: 1393
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20040048253A1 LI:351965.1:2000FEB01
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 120, 765
;   OTHER INFORMATION: a, t, c, g, or other
US-10-220-120-17

  Query Match             58.6%;  Score 778.8;  DB 16;  Length 1393;
  Best Local Similarity   87.2%;  Pred. No. 1.3e-231;
  Matches  904;  Conservative    0;  Mismatches  112;  Indels   21;  Gaps    4;

Qy        302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAA 361
              ||||||||||| |||| ||| |||||||||| ||| |||||||||||||||||||||||
Db          1 GAAGCTGTAAAACTACTAATTGTTCTTCAAAAGGAGCCCCAGATGCAGTCCATGACTAAG 60

Qy        362 ATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
              |||||||   |||||||||||||||||||||| | |||| | ||||||||||| || ||
Db         61 ATCTACCACAGACCCCTGGACCGGCCTGCTAGCCCATGCACCGATGTTAATGATATCGAN 120

Qy        422 GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGA 481
                ||| || ||| ||||||| |||||||||||||||||||||||| ||||||||| ||||
Db        121 AGCACTCC-CCCAAGGAAATTTCAACTGCACAACCCCTACTACACCCCAATTCAGCAGGA 179

Qy        482 AGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAG 541
              |||||||||||| || ||||||||||||||||| ||| ||||||||||||||||||||||
Db        180 AGCAGTTAGAGCGGTCGTCAGCCAACCTCCCCAGCAGCACTTGGGTTTTCCTGTTGAGAG 239

Qy        542 GGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTA 601
               | ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        240 CGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTA 299

Qy        602 NCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGA 661
               |||||||||||||||| || | ||||||||| |||||||||||||||||||||||||| |
Db        300 GCTGGGAAGGTGACCGCGTCTACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAA 359

Qy        662 CCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAG 712
              ||||||         ||||||||||||||||||||||| ||||||||||||||||||| ||
Db        360 CCAATCAGGTAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAGAG 419

Qy        713 CAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAA 772
               ||||||||||||||||| ||||||||||||||| ||| |||||| |||   |||||||||
Db        420 AAATAGCCAATCATCTATCGCCTGAGAGCACAGCAGGAGGGACAATGATCCGGATATAAA 479

Qy        773 CTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGC 832
              | || ||||||| |||||||||| ||  |||||||||| |||||||||| ||||||||||||
Db        480 CCCAAGCATTCGAGCCAGCAACGGCTGCCCCCTTTGTGTCCCCTCCCTTTGTATGGGAGC 539

Qy        833 TCTGTTTTCACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTT 892
              |||||          ||||||||||||||||| ||||||||||||||||||||||| |||||
Db        540 TCTGT----------TTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCTGTGTT 589

Qy        893 TTTTATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACA 952
              | ||| || |  |||||||||||| | ||||| |||||||||||||||||||| || ||  |
Db        590 TGTTACGGTTTGAGCTGAGCTTTCGCTCGCCGTCCACCACTGCTGTTTGCCGCCATCGGA 649

Qy        953 GACCCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCC 1012
              |||| || ||||||||||||||||  |||||  |||| |||| || ||||||||||||||
Db        650 GACCTGCCGCTGACTTCCATCCCTCCGGATCTGGCAGGGTGTTCA-TGTGCTCCTGATCC 708

Qy       1013 AGCGAGGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGC 1072
              || |||| ||||||||||| ||| |||  |||||||||||||||||||||||||| | |
Db        709 AGAGAGGCACCCATTGCCATTCCTGATTGGGCTAAAGGCTTGCCATTGTTCCTGCANGAC 768

Qy       1073 TAAGTGCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCT 1132
              |||||||| ||||  |||||||| || |||||||||| ||| ||||||||||| | ||||||
Db        769 TAAGTGCCCGGGTTCATCCTAATCGAGCTGAACACTAGTCGCTGGGTTCCACGATTCTCT 828

Qy       1133 TCCATGACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATT 1192
              ||| |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||
Db        829 TCCGTGACCCACGACTTCTAATAGAGCTATAACACTCACCGCACGGCCCAAGATTCCATT 888

Qy       1193 CCTTGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTG 1252
              |||| | ||| ||||||||||||||||||||||||||||  |||||||||||||||||||
Db        889 CCTTCGAATCCGTGAGGCCAAGAACCCCAGGTCAGAGAACATGAGGCTTGCCACCATTTT 948

Qy       1253 GGAAGTGGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCA 1312
              ||||||||||  | |||||||||| || ||||  |||||||||||||||||| ||||||||
Db        949 GGAAGTGGCCTGCCGCCATTTTGGAAGTGGCCTGCCACCATCTTGGGAGCTCTGGGAGCA 1008

Qy       1313 AGGATCCCCCAGTAACA 1329
              |||| ||||  ||||||
Db       1009 AGGACCCCCTGGTAACA 1025
```

```
RESULT 6
US-10-637-565-18
; Sequence 18, Application US/10637565
; Publication No. US20040043381A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  PERRON, Herve
;  APPLICANT:  KOMURIAN-PRADEL, Florence
;  TITLE OF INVENTION: THE LTR REGION OF MSRV-1 AND THE PROTEINS IT ENCODES, AND PROBES
;  TITLE OF INVENTION:  METHODS FOR DETECTING MSRV-1 RETROVIRUS
;  FILE REFERENCE: 110257
;  CURRENT APPLICATION NUMBER: US/10/637,565
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: US/09/890,340
;  PRIOR FILING DATE: 2001-11-27
;  PRIOR APPLICATION NUMBER: PCT/IB00/00159
;  PRIOR FILING DATE: 2000-02-15
;  PRIOR APPLICATION NUMBER: EP 99420041.8
;  PRIOR FILING DATE: 1999-02-15
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 2030
;   TYPE: DNA
;   ORGANISM: MSRV-1 retrovirus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1626)
US-10-637-565-18

  Query Match             58.3%;  Score 775.4;  DB 16;  Length 2030;
  Best Local Similarity   92.8%;  Pred. No. 1.9e-230;
  Matches  812;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 1199

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              ||| |||||| ||||||| |||||||||||||||  ||||  | ||||||||||||||||||
Db       1200 AGGAGAAGAACGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGA 1259

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| |||||||||||||||| |||||||||| | || |||||||||
Db       1260 AATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCT 1319

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||  |||||||||||||||||||||||||||| |||||||
Db       1320 CCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATT 1379

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               |||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||
Db       1380 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAAT 1439

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
              |||||||||||||||||| || ||  | ||||||||||||||||||| ||||||||||||||
Db       1440 TGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCAGATGGAGTCCATGACTAA 1499

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||| |||||||||||||||||||||||||||| | |||||| ||||||||||||||||||
Db       1500 GATCCACCGTGGACCCCTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGA 1559

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||||||||||||||||||||||||||||||||||  | |||||||||  ||
Db       1560 AGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGG 1619

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| ||  |||||||||||||||||||| ||||||||||||||||||||||
Db       1620 AAGCAGTTAGAGCGGTCATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGA 1679

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||| ||||||||||||||||||||||||||||||||||||| || |||||||||| ||||||
Db       1680 GGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCT 1739

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | ||||||||||||| |||||||| || |||||||||||||||||||||||||||||||||
Db       1740 AGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCG 1799

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||| |||||||||||| |||||||
Db       1800 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCC 1859

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||||||| ||||||||||| ||||||||||||| ||||||
Db       1860 AATCATCTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCA 1919

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||||| ||||||||| ||||||||||
Db       1920 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTT 1979

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
```

```
               |||||||||||||||||||||||| |||||||| |
Db        1980 CACTCTATTTCACTCTATTAAATCTTGCAACTGAA 2014

RESULT 7
US-10-198-846-9936
; Sequence 9936, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9936
;   LENGTH: 7974
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-198-846-9936

  Query Match             58.3%;  Score 774.2;  DB 14;  Length 7974;
  Best Local Similarity   85.8%;  Pred. No. 9.2e-230;
  Matches  937;  Conservative    0;  Mismatches  126;  Indels   29;  Gaps    6;

Qy         257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
               ||||||||||| ||| ||||||||||||||||||||||||||||  |||||||||| |||
Db        6409 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGCTGTAAAACTA 6468

Qy         317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGAC-TAAAATCTACCGTGGACC 375
               ||||| |||||||||||||| |||||||||||||||||||| ||| ||||||| ||||||
Db        6469 CAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTTAAGATCTACCATGGACC 6528

Qy         376 CCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGA 435
               |||||||| ||||||||| | ||||||||||||||||||||| |||| ||||||||| ||
Db        6529 CCTGGACCAGCCTGCTAGCCCATGCTCTGATGTTAATGACATCGAAGACACCCCTCCAGA 6588

Qy         436 GGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAG 495
               ||||||||||||||||||||||||||||  | |  |||||| |||||| ||||||||  |
Db        6589 GGAAATCTCAACTGCACAACCCCTACTATGCCCTGATTCAGCAGGAAGTAGTTAGAGTGG 6648

Qy         496 TTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGA 555
               ||||| ||||||||||||||||| ||||| |||||||||||| ||||| |||||||||||
Db        6649 TTGTCGGCCAACCTCCCCAACAGCACTTGAGTTTTCCTGTTGGGAGGGGGGACTGAGAGA 6708

Qy         556 CAGGACTAGCTGGATTTCCTA-------GGCTGACTAAGAATCCCNAAGCCTANCTGGGA 608
               |||||||||||||  || |         | || |  ||  | | | ||||||| ||||||
Db        6709 CAGGACTAGCTGGGATTTCCCTAGGGCCGACTTAAGAATCACCACTAAGCCTAGCTGGGA 6768

Qy         609 AGGTGACCGCAT-CCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC 667
               |||||||||| | ||| ||||||||| |||||||||||||||||||||||||||||||||
Db        6769 AGGTGACCGCGTCCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCGACCAATC 6828

Qy         668 ---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAG 718
                        |||||||||  ||||||| |||| ||||||||||||||||||||| |||||
Db        6829 AGGTAGCAAAGAGAGCTCGTTAAAATGATAATTAGGCAAAAACAGGAGGTAAAGAAATAG 6888

Qy         719 CCAATCATC-TATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAG 777
               ||||||||| |||  ||||||||||||  ||||||||||| ||| ||||||||||| |||
Db        6889 CCAATCATCTTATCACCTGAGAGCACAATGGGAAGGACAATGATCGGGATATAAACCCAG 6948

Qy         778 GCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGT 837
               |||||| ||   ||||| || |||| |||||||||||||||| |||||||| ||||||||
Db        6949 GCATTCCAGTTGGCAACTGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGAAGCTCTGT 7008
```

```
Qy         838 TTTCACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTA 897
                         ||||||||||||||||| |||  ||||| ||||||||||| | ||| |||
Db        7009 ----------TTTCACTCTATTAAATCTTGCTTCTGCAGTCTTCTGGTCCATATTTGTTA 7058

Qy         898 TGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCC 957
                ||||| ||||||||||||| ||||| |||||| ||| |||||||| | ||| |||||||
Db        7059 CGGCTCCAGCTGAGCTTTTGCTCGCCGTCCACCCCTGTTGTTTGCCGCTGTCGCAGACCC 7118

Qy         958 GCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGA 1017
               || ||||||||||||||||  ||||||||||| |||||  ||||||||||||| ||||||
Db        7119 GCCGCTGACTTCCATCCCTCCGGATCCAGCAGGGTGTCTGCTGTGCTCCTGATGCAGCGA 7178

Qy        1018 GGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGT 1077
               ||  |||||| ||||||| |||| | || |||||||||||||||||||||| |||||||
Db        7179 GGCGCCCATTTCCACTCCTGATCGGACTGAAGGCTTGCCATTGTTCCTGCACGGCTAAGA 7238

Qy        1078 GCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCAT 1137
               ||||||||| |||||||| || ||||||||| |||||||||||||| ||||||||||| |
Db        7239 GCCTGGGTTCGTCCTAATCGAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCCGT 7298

Qy        1138 GACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTG 1197
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7299 GACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTG 7358

Qy        1198 GTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
               | ||| ||||||||||||||||||||||||||||   ||||||||||||| | | |||||
Db        7359 GAATCCGTGAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCGTCTTGGAAG 7418

Qy        1258 TGGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGAT 1317
               |||||  |  |||| |||| |||||||  ||||||||||||||||| ||||||||||||
Db        7419 TGGCCTGCCACCATCTTGGAAGCGGCCTGCCACCATCTTGGGAGCTCTGGGAGCAAGGAC 7478

Qy        1318 CCCCCAGTAACA 1329
               ||||  ||||||
Db        7479 CCCCTGGTAACA 7490

RESULT 8
US-10-416-642-4
; Sequence 4, Application US/10416642
; Publication No. US20040043452A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  RAMKUMAR, Jayalaxmi
;  APPLICANT:  ARVIZU, Chandra
;  TITLE OF INVENTION: EMBRYOGENESIS ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-0842 PCT
;  CURRENT APPLICATION NUMBER: US/10/416,642
;  CURRENT FILING DATE:  2003-05-13
;  PRIOR APPLICATION NUMBER: 60/249,407
;  PRIOR FILING DATE: 2000-11-15
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PERL Program
; SEQ ID NO 4
;   LENGTH: 2074
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20040043452A1 7477736CB1
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1995
;   OTHER INFORMATION: a, t, c, g, or other
US-10-416-642-4

  Query Match             57.7%;  Score 766.8;  DB 16;  Length 2074;
  Best Local Similarity   93.0%;  Pred. No. 9.1e-228;
  Matches  812;  Conservative    0;  Mismatches   60;  Indels    1;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1203 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 1262
```

```
Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               ||| |||||||||||||| |||||||||||||||  |||||  | |||||||||||||||||
Db        1263 AGGAGAAGAATGCTGTTATTATGTTAATCAATCCAGAATTGTCACTGAGAAAGTTAAAGA 1322

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| |||||||||||| ||||||||||||||||| |||||||||||| | || |||||||||
Db        1323 AATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAACACTGGGGCCT 1382

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||  |||||||||||||||||||||||||||||| |||||||
Db        1383 CCTCAGCCAATGGATGCCCTGGGTTCTCCCCTTCTTAGGACCTCTAGCAGCTCTAATATT 1442

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||
Db        1443 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAAT 1502

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               |||||||||||||||||||||| |||||||||||||| |||||||||||||||||||||||
Db        1503 TGAAGCTGTAAAGCTACAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAA 1562

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||||| | |||||| ||||||||||||||| ||
Db        1563 AATCTACCACGGACCCCTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGA 1622

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||| ||||||| ||||||||||||||||||||||||||||||  | |||||||||| |||
Db        1623 AGGCACTCCTCCCAAGGAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCAGG 1682

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||||| ||||||||||||||||| ||||||||||||||||||||
Db        1683 AAGCAGTTAGAGCGGTCGTCAGTCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGA 1742

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               ||| |||||||||||||||||||||||||||||||||||| || |||||||||| ||||||
Db        1743 GGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGATTAAGAATCCCTAAGCCT 1802

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | ||||||||||||||||| ||| |||||||||| |||||||||||||||||||||||||||
Db        1803 AGCTGGGAAGGTGACCGCGTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCA 1862

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db        1863 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 1922

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               ||||||||||||||||||||||||| |||| |||||||||||||  ||||||||| ||||||
Db        1923 AATCATCTATTGCCTGAGAGCACAGTGGGAGGGACAAGGATTGCAATATAAACCCAGGCA 1982

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| ||||||| | ||||||| |||||||||||| |||| ||||||||||||||||||||
Db        1983 TTCGAGCCAGC-ANAGCAACCGCCTTTGGGTCCCTTCCCCTTGTATGGGAGCTCTGTTTT 2041

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTG 873
               |||||||||||||||||||||||| ||||||||
Db        2042 CACTCTATTTCACTCTATTAAATCTTGCAACTG 2074

RESULT 9
US-10-114-893-134
; Sequence 134, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT: McCoy, John M.
;  APPLICANT: LaVallie, Edward R.
;  APPLICANT: Collins-Racie, Lisa A.
;  APPLICANT: Evans, Cheryl
;  APPLICANT: Merberg, David
;  APPLICANT: Treacy, Maurice
;  APPLICANT: Bowman, Michael R.
;  APPLICANT: Spaulding, Vikki
;  APPLICANT: Carlin-Duckett, McKeough
;  APPLICANT: Kelleher, Kerry S.
;  APPLICANT: Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: GI 6000-10A
;  CURRENT APPLICATION NUMBER: US/10/114,893
;  CURRENT FILING DATE: 2002-04-02
;  EARLIER APPLICATION NUMBER: 09/413,232
;  EARLIER FILING DATE: 1999-10-06
;  NUMBER OF SEQ ID NOS: 321
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 134
;   LENGTH: 2946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-114-893-134

  Query Match             55.1%;  Score 731.8;  DB 13;  Length 2946;
  Best Local Similarity   90.7%;  Pred. No. 9.1e-217;
  Matches  794;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        2067 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 2126

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        2127 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2186

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| |||||||||||| || |||||||||||||| |||| |||||||| ||||||||||||
Db        2187 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2246

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        2247 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2306

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        2307 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2366

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| ||||||||||||| ||||||
Db        2367 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2414

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        2415 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2474

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || |||||||||| |||||||||||| |||||||||| ||||||||| | |||||||| |||
Db        2475 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2534

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||||| |||| ||||||||||||||||
Db        2535 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2594

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        2595 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2654

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| |||||||||| |||||||||||||||||||||||| |
Db        2655 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2714

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db        2715 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2774

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| |||||||||||| || |
Db        2775 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2834

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
```

```
              ||| |||| ||||| ||||||||||||||||||||||| |||||||||||||||||||||
Db        2835 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTT 2894

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db        2895 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2929

RESULT 10
US-10-016-249-3
; Sequence 3, Application US/10016249
; Publication No. US20030100053A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Mi, Sha
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6006B.AJ172A
;  CURRENT APPLICATION NUMBER: US/10/016,249
;  CURRENT FILING DATE:  2001-10-30
;  PRIOR APPLICATION NUMBER: US/09/175,928
;  PRIOR FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-016-249-3

  Query Match             55.1%;  Score 731.8;  DB 15;  Length 2946;
  Best Local Similarity   90.7%;  Pred. No. 9.1e-217;
  Matches  794;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        2067 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 2126

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        2127 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2186

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        2187 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2246

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        2247 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2306

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        2307 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2366

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        2367 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2414

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        2415 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2474

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        2475 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2534

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
```

```
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        2535 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2594

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        2595 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2654

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2655 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2714

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db        2715 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2774

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        2775 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2834

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||| |||| ||||| ||||||||||||||||||||||| |||||||||||||||||||||
Db        2835 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTT 2894

Qy         841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db        2895 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2929

RESULT 11
US-09-902-535-1
; Sequence 1, Application US/09902535
; Patent No. US20020102530A1
; GENERAL INFORMATION:
;  APPLICANT: Keith, Jr., James C.
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  Mi, Sha
;  TITLE OF INVENTION: Methods and compositions for diagnosing
;  TITLE OF INVENTION:  and treating preeclampsia and gestational trophoblast
;  TITLE OF INVENTION:  disorders
;  FILE REFERENCE: GIN-6006B4
;  CURRENT APPLICATION NUMBER: US/09/902,535
;  CURRENT FILING DATE:  2001-07-09
;  PRIOR APPLICATION NUMBER: 60/216,657
;  PRIOR FILING DATE: 2000-07-06
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2930
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (930)...(2546)
US-09-902-535-1

  Query Match             55.0%;  Score 730.8;  DB 9;  Length 2930;
  Best Local Similarity   90.7%;  Pred. No. 1.9e-216;
  Matches  793;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        2069 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 2128

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        2129 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2188

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        2189 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2248

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2249 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2308

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db       2309 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2368

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db       2369 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2416

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db       2417 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2476

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2477 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2536

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2537 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2596

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       2597 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2656

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2657 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2716

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db       2717 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCC 2776

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2777 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2836

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| |||||||||||||||||||
Db       2837 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTTT 2896

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGC 874
              ||  |||||||||||||||||||| |||||||||
Db       2897 CATGCTATTTCACTCTATTAAATCTTGCAACTGC 2930

RESULT 12
US-10-632-793-30
; Sequence 30, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
;   LENGTH: 7582
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (307)..(307)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (355)..(355)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1309)..(1309)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1331)..(1331)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2213)..(2213)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2398)..(2398)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3787)..(3787)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4115)..(4115)
;   OTHER INFORMATION: n = a or g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4261)..(4261)
;   OTHER INFORMATION: n = a or g or c or t/u
US-10-632-793-30

  Query Match             54.4%;  Score 722.6;  DB 16;  Length 7582;
  Best Local Similarity   88.5%;  Pred. No. 1.1e-213;
  Matches  774;  Conservative   19;  Mismatches   70;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||:||||||||||||||||||||:|||||||||  |:|||||||||||||||||||||||
Db       6720 TCRAAATCGAAGAGCTTTAGACTYGCTAACCGCTGARAGAGGGGGAACCTGTTTATTTTT 6779

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | ||||||||||||:||||
Db       6780 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTRAAGA 6839

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| :|||||||||| || |||:||||||:|| |||| |||||||| ||||||||||||
Db       6840 AATTCSAGATCGAATACAACGTAKAGCAGARGAGCTTCGAAACACTGGACCCTGGGGCCT 6899

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||:|||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       6900 CCTCAGCCRATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 6959

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| :|| |||||||||| |||||||||||||||||
Db       6960 GCTACTCCTCTTTGGACCCTGTATCTTTRACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 7019

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||:|| |||            |||||||| ||| |||||||||||| ||||||
Db       7020 CGAAGCTGTRAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 7067

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  ||||||||||||||||||:||| | | | |||||||||||||||||  |
Db       7068 GATCTACCGCAGACCCCTGGACCGGCCTGYTAGCCCACGATCTGATGTTAATGACATCAA 7127
```

```
Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       7128 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 7187

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       7188 AAGCAGTTAGAGCGGTSGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 7247

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||:|||||||||: ||||||
Db       7248 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGAYTAAGAATCCYTAAGCCT 7307

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | :||||||||||||| |||||| ||||||||| ||||||||||||||||:||||||| |
Db       7308 AGSTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGYTCACACCTG 7367

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       7368 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 7427

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||:||||||:|||||||||||| ||| |||||| ||| ||||||||||| || |
Db       7428 AATCATYTATTGCMTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTY 7487

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       7488 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 7547

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||:||||:
Db       7548 CATGCTATTTCACTCTATTAAATCTTGCARCTGCR 7582

RESULT 13
US-10-632-793-25
; Sequence 25, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 1136
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-632-793-25

  Query Match             54.3%;  Score 721.2;  DB 16;  Length 1136;
  Best Local Similarity   90.0%;  Pred. No. 1.1e-213;
  Matches  787;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||||||||||||||||||| |||||||||  |||||||||||||||||||||||||
Db        254 TCAAAATCGAAGAGCTTTAGACTCGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 313

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        314 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 373

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || |||||||||| || |||| |||||||| ||||||||||||
Db        374 AATTCGAGATCGAATACAACGTAGAGCAGAAGAGCTTCGAAACACTGGACCCTGGGGCCT 433

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        434 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 493

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        494 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 553

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        554 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 601

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        602 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 661

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        662 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 721

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        722 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 781

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        782 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 841

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||| ||||||| |
Db        842 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGTTCACACCTG 901

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        902 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 961

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              ||||||||||||| |||||||||||| ||| |||||| ||| ||||||||||| || |
Db        962 AATCATCTATTGCATGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 1021

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       1022 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 1081

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGC 874
              ||  |||||||||||||||||||| |||| ||||
Db       1082 CATGCTATTTCACTCTATTAAATCTTGCAGCTGC 1115

RESULT 14
US-10-632-793-26
; Sequence 26, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
;  APPLICANT: PARANHOS-BACCALA, Glaucia
;  APPLICANT:  MALLET, Francois
;  APPLICANT:  VOISSET, Cecile
;  TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
;  TITLE OF INVENTION:  AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
;  FILE REFERENCE: 110048
;  CURRENT APPLICATION NUMBER: US/10/632,793
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: US/09/869,927
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/FR00/00144
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: FR 99/00888
;  PRIOR FILING DATE: 1999-01-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
;   LENGTH: 2782
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-632-793-26

  Query Match             53.6%;  Score 712.6;  DB 16;  Length 2782;
  Best Local Similarity   89.4%;  Pred. No. 8.7e-211;
  Matches  782;  Conservative    0;  Mismatches   81;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              || ||||||||||||||||||||||||||||||  | |||||||||||||||||||||||
Db       1902 TCGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTT 1961

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||| ||||
Db       1962 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGA 2021

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  |||||||||| || ||| ||||||||| |||| |||||||| ||||||||||||
Db       2022 AATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2081

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              |||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2082 CCTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2141

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||||||||||||||||||||||||  || |||||||||| |||||||||||||||||
Db       2142 GCTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2201

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               |||||||| || |||            |||||||| ||| |||||||||||| ||||||
Db       2202 CGAAGCTGTGAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2249

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | |||||||||||||||||  |
Db       2250 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2309

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2310 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2369

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2370 AAGCAGTTAGAGCGGTGGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2429

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db       2430 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCT 2489

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              |  ||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2490 AGGTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2549

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2550 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2609

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||| ||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2610 AATCATTTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTT 2669

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTT 840
              ||| |||| ||||| ||||||||||||||||||||||| |||||||||||||||||||||
Db       2670 TTCGAGCCGGCAACGGCAACCCCCTTTGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTT 2729

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2730 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2764

RESULT 15
US-10-133-036-1
```

```
; Sequence 1, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
;  APPLICANT: Conrad, Bernard
;  APPLICANT:  Mach, Bernard
;  TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
;  FILE REFERENCE: 23135-507
;  CURRENT APPLICATION NUMBER: US/10/133,036
;  CURRENT FILING DATE:  2002-04-26
;  PRIOR APPLICATION NUMBER: PCT/EP00/10659
;  PRIOR FILING DATE: 2000-10-30
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2782
;   TYPE: DNA
;   ORGANISM: Human endogenous retrovirus
US-10-133-036-1

  Query Match             53.5%;  Score 711;  DB 16;  Length 2782;
  Best Local Similarity   89.3%;  Pred. No. 2.7e-210;
  Matches  781;  Conservative    0;  Mismatches   82;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              || ||||||||||||||||||||||||||||||  | |||||||||||||||||||||||
Db       1902 TCGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTTATTTTT 1961

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||| ||||
Db       1962 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGA 2021

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  |||||||||| || ||| ||||||||| |||| |||||||| ||||||||||||
Db       2022 AATTCCAGATCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2081

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              |||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2082 CCTCAGCCGATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2141

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||||||||||||||||||||||||  || |||||||||| |||||||||||||||||
Db       2142 GCTACTCCTCTTTGGACCCTGTATCTTTGACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2201

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               |||||||| || |||            |||||||| ||| |||||||||||| ||||||
Db       2202 CGAAGCTGTGAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2249

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               |||||| |  |||||||||||||||||||||| | | |||||||||||||||||  |
Db       2250 GATCTACGGAAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2309

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db       2310 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2369

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db       2370 AAGCAGTTAGAGCGGTGGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2429

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db       2430 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCT 2489

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              |  ||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db       2490 AGGTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2549

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db       2550 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2609

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||| ||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db       2610 AATCATTTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTT 2669

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db       2670 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2729

Qy        841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
              ||  |||||||||||||||||||| ||||||||||
Db       2730 CATGCTATTTCACTCTATTAAATCTTGCAACTGCA 2764


RESULT 16
US-09-864-761-4444
; Sequence 4444, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4444
;   LENGTH: 1894
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC002346.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =   6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444

  Query Match             53.0%;  Score 704.8;  DB 9;  Length 1894;
  Best Local Similarity   89.7%;  Pred. No. 1.9e-208;
  Matches  770;  Conservative    0;  Mismatches   79;  Indels    9;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||| |||||||||| |||||||||
Db        961 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGGGAACCTTTTTATTTTT 1020

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              || |||| |||||||||  |||||||||||||| |||||||| || ||||||||||||||
Db       1021 AGAGGAAAAATGCTGTTGTTATGTTAATCAATCCGGAATCATCACCGAGAAAGTTAAAGA 1080

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              ||||  || |||||||||| |||||||| |||| || |||||||||| ||||||||||||
Db       1081 AATTCAAGGTCGAATATAACGTAGAGCAAAGGAGCTGCAAAACACTGGACCCTGGGGCCT 1140

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1141 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1200

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
               |||||||||||||||||||||||||| ||| |||||||||||||||||| |||||||||
Db       1201 GTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAAT 1260

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||| ||||| |||||||| |||||||||||||| ||||||||||||||||||| |||
Db       1261 CGAAGCAGTAAAACTACAAATCGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGAGTAA 1320

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
              ||||||||  ||||||||||||||||||||||| | |||||||||||||||||||||  |
Db       1321 AATCTACCACGGACCCCTGGACCGGCCTGCTAGCCCATGCTCTGATGTTAATGACATCAA 1380

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||||||||||||||||||||||||||| ||||||| | ||||||||| |||
Db       1381 AGGCACCCCTCCCGAGGAAATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGG 1440

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||  |||||  ||||||||||||||||| | |||||||||||||||||||
Db       1441 AAGCAGTTAGAGTGGTTGTTGGCCAACCTCCCCAACAGCAGTTGGGTTTTCCTGTTGAGA 1500

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
              ||| |||||||||||||||| || || |||||||||| |  |||||||||||| ||| ||
Db       1501 GGGGGGACTGAGAGACAGGAATAACTAGATTTCCTAGACCAACTAAGAATCCCTAAGACT 1560

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||||| |||| |||||||||  |||||||||||||||||||| |||
Db       1561 AGCTGGGAAGGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCA 1620

Qy        661 ACCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAA 711
              |||||||         ||||||||||||||||||||||| ||||||||||||||| ||||
Db       1621 ACCAATCAGATACTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGATAAA 1680

Qy        712 GCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAA 771
              | ||||||||||||||| |||||||| |||||||| ||| |||||| ||| |||||||||
Db       1681 GAAATAGCCAATCATCTGTTGCCTGACAGCACAGCAGGAGGGACAATGATCGGGATATAA 1740

Qy        772 ACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAG 831
              || ||||||||| ||||||| ||||| |||| |||||||||||||||| |||||||||||
Db       1741 ACCCAGGCATTCGAGCCAGCTACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAG 1800

Qy        832 CTCTGTTTTCACTCTATT 849
              |||||| |||||||||||
Db       1801 CTCTGTCTTCACTCTATT 1818
```

```
RESULT 17
US-10-087-192-730/c
; Sequence 730, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 730
;   LENGTH: 161334
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(161334)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-730

  Query Match             48.4%;  Score 643.4;  DB 13;  Length 161334;
  Best Local Similarity   85.4%;  Pred. No. 3e-188;
  Matches  781;  Conservative    0;  Mismatches   88;  Indels   46;  Gaps    4;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              ||||||| ||||||||||||||||||||||| |  |||||||||||| ||||||||||||
Db     100898 TCAAAATTGAAGAGCTTTAGACTTGCTAACCTCTGAAAGAGGGGGAAGCTGTTTATTTTT 100839

Qy         61 AGGGGAAGAATGCTGTTAG------------------------------TATGTTAATCA 90
              |||||||||||| |||||                              |||||||||||
Db     100838 AGGGGAAGAATGTTGTTATTATGTTATTTTAGCGGAAGAATGTTGTTATTATGTTAATCA 100779

Qy         91 ATCTGGAATCATTACTGAGAAAGTTAAAGAAATTTGAGATCGAATATAATGTAGAGCAGA 150
              |||  ||||  | || ||||||||| |||||||| ||||| ||||| || ||||| ||||
Db     100778 ATCCTGAATTGTCACAGAGAAAGTTGAAGAAATTCGAGATTGAATACAACGTAGAACAGA 100719

Qy        151 GGACCTTC-AAAACACTGCACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGACTCTCC 209
              ||| |||| |||||||   ||||||||||||||||||||||||||||||||||| |||||
Db     100718 GGAGCTTCAAAAACACCAGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCC 100659

Qy        210 CCTTCTTAGGACCTCTAGCAGCTATAATATTTTTACTCCTCTTTGGACCCTGTATCTTCA 269
              ||||||||||| ||||||||||| |||||||  ||||||||||||||||||||||||| |
Db     100658 CCTTCTTAGGATCTCTAGCAGCTCTAATATTGATACTCCTCTTTGGACCCTGTATCTTTA 100599

Qy        270 ACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAAATAGTTCTTC 329
              || ||||||||||||||||||||||||||||  ||| ||||||||||||||| |||||||
Db     100598 ACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGTTGTAAAGCTACAAATCGTTCTTC 100539

Qy        330 AAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCCCTGGACCGGCCTG 389
              ||||||||||||||||| ||||||||||||| |||||||||||||||||||||||||||
Db     100538 AAATGGAACCCCAGATGAAGTCCATGACTAAGATCTACCGTGGACCCCTGGACCGGCCTA 100479

Qy        390 CTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAGGAAATCTCAACTG 449
              |||| | ||||||  ||  |||||| || |||  ||||||||||||||||||||||||||
Db     100478 CTAGCCCATGCTCCAATTGTAATGATATCGAACGCACCCCTCCCGAGGAAATCTCAACTG 100419

Qy        450 CACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGTTGTCAGCCAACCT 509
              ||||||||||||||  | ||||||| | ||||||||||||||   || ||||||||||||
Db     100418 CACAACCCCTACTATGCCCCAATTCCGCAGGAAGCAGTTAGACTGGTCGTCAGCCAACCT 100359

Qy        510 CCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGACAGGACTAGCTGGA 569
              ||||||||| |||||||||||||||||||| ||| |||||||||||||||| ||||||||
Db     100358 CCCCAACAGCACTTGGGTTTTCCTGTTGAGTGGGGGGACTGAGAGACAGGATTAGCTGGA 100299

Qy        570 TTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACCGCATCCATCTTTA 629
              |||||||||| ||||||||||||| ||||||| |||||||||||||| |||||| |||||
Db     100298 TTTCCTAGGCCGACTAAGAATCCCAAAGCCTAGCTGGGAAGGTGACCACATCCACCTTTA 100239

Qy        630 AACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC---------AGAGAGCTCACTA 680
              ||||  ||||||||||||||||||||||||||||||||         |||||||||||||
Db     100238 AACACTGGGCTTGCAACTTAGCTCACACCCGACCAATCAGGTAGTAAAGAGAGCTCACTA 100179

Qy        681 AAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTGAGAG 740
              |||||||||| || |||||||||||||||||||||||||||  ||||||| |||||||||
Db     100178 AAATGCTAATTAGACAAAAACAGGAGGTAAAAAAATAGCCAATCATCTATCGCCTGAGAG 100119

Qy        741 CACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAGCAAC 800
              ||||||||||||||||| ||| ||||||||||| |||||||||||||| ||||| || ||
Db     100118 CACAGCGGGAAGGACAATGATCGGGATATAAACCCAGGCATTCAAGCCGGCAACGGCTAC 100059

Qy        801 CCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCTATTA 860
              |  |||||||||||||||| ||||||||||||||| | |         ||||||||||||
Db     100058 CTTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTCTCT------GTCTTCACTCTATTA 100005

Qy        861 AATCATGCAACTGCA 875
              |||  ||||||||||
Db     100004 AATATTGCAACTGCA 99990

RESULT 18
US-09-854-867-385
; Sequence 385, Application US/09854867
; Publication No. US20030224356A1
; GENERAL INFORMATION:
;  APPLICANT: JOAN, KNOLL H
;  APPLICANT:  ROGAN, PETER K
;  TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
;  FILE REFERENCE: 30307
;  CURRENT APPLICATION NUMBER: US/09/854,867
;  CURRENT FILING DATE:  2003-05-08
;  NUMBER OF SEQ ID NOS: 613
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 385
;   LENGTH: 780
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: rerpeat_region
;   LOCATION: (1)..(780)
;   OTHER INFORMATION: ltr17
US-09-854-867-385

  Query Match             48.0%;  Score 637.8;  DB 10;  Length 780;
  Best Local Similarity   90.0%;  Pred. No. 9.5e-188;
  Matches  711;  Conservative    0;  Mismatches   60;  Indels   19;  Gaps    2;

Qy        549 TGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGA 608
              ||||||||||||||||||||||||||||||| ||||||||||||| ||||||| ||||||
Db          1 TGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGGA 60

Qy        609 AGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC- 667
              |||||||||| |||| ||||||||| |||||||||||||||||||||||||||||||||
Db         61 AGGTGACCGCTTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCGACCAATCA 120

Qy        668 --------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGC 719
                      ||||||||||||||||||||||| ||||||||||||||||||||| ||||||
Db        121 GGTAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGC 180

Qy        720 CAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGC 779
              ||||||||||| ||||||||||||||||||| |||||| ||| ||||||||||| |||||
Db        181 CAATCATCTATCGCCTGAGAGCACAGCGGGAGGGACAATGATCGGGATATAAACCCAGGC 240

Qy        780 ATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTT 839
              |||| |||| ||||| || |||| |||||||||||||||| |||||||||||||||||
Db        241 ATTCGAGCCGGCAACGGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT-- 298

Qy        840 TCACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATG 899
                      ||||||||||||||||| |||||||||||||||||||||||||||| ||| |
Db        299 --------TTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCGTGTTTGTTACG 350

Qy        900 GCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGC 959
              |||| ||||||||||| | || || ||||||||||||||||||| ||||| |||||||||
Db        351 GCTCGAGCTGAGCTTTCGCTCACCGTCCACCACTGCTGTTTGCCGCCGTCGCAGACCCGC 410

Qy        960 TGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGG 1019
               ||||||||||||||||  |||||| |||| |||||| ||||||||||||||||||||||
Db        411 CGCTGACTTCCATCCCTCCGGATCCGGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGG 470

Qy       1020 TACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGC 1079
                ||||||||| ||||||||| ||||||||||||||||||||||||||| ||||||||||
Db        471 CGCCCATTGCCGCTCCCGATCGGGCTAAAGGCTTGCCATTGTTCCTGCACGGCTAAGTGC 530

Qy       1080 CTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGA 1139
              ||||||| |||||||| || ||||||||| |||||||||||||| ||||||||||| |||
Db        531 CTGGGTTCGTCCTAATCGAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCCGTGA 590

Qy       1140 CCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGT 1199
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        591 CCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGA 650

Qy       1200 ATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTG 1259
              ||| ||||||||||||||||||||||||||||   ||||||||||||||| | |||||||
Db        651 ATCCGTGAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGAAGTG 710

Qy       1260 GCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCC 1319
              |||| | |||||||||| |||||||| ||||||||||||||||| |||||||||||| ||
Db        711 GCCCGCCGCCATTTTGGAAGCGGCCCGCCACCATCTTGGGAGCTCTGGGAGCAAGGAACC 770

Qy       1320 CCCAGTAACA 1329
              ||| ||||||
Db        771 CCCGGTAACA 780

RESULT 19
US-10-672-764A-34/c
; Sequence 34, Application US/10672764A
; Publication No. US20040156832A1
; GENERAL INFORMATION:
;  APPLICANT: Jolly, Chris
;  TITLE OF INVENTION: Immunoglobulin Compositions and Methods
;  FILE REFERENCE: 13311.1001U
;  CURRENT APPLICATION NUMBER: US/10/672,764A
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 68
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
;   LENGTH: 157090
;   TYPE: DNA
;   ORGANISM: Human
US-10-672-764A-34

  Query Match             45.3%;  Score 601.4;  DB 17;  Length 157090;
  Best Local Similarity   87.0%;  Pred. No. 3.8e-175;
  Matches  689;  Conservative    0;  Mismatches   84;  Indels   19;  Gaps    2;

Qy        547 ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGG 606
              | ||||||||||||||||||||||||||||||| ||||||||||||| ||||| | ||||
Db      81195 ATTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCCAGCTGG 81136

Qy        607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 666
              |||||||||| | |||| ||||||||| ||||||||||||| ||||||||||||||||||
Db      81135 GAAGGTGACCACTTCCACCTTTAAACACGGGGCTTGCAACTCAGCTCACACCCGACCAAT 81076

Qy        667 C---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATA 717
              |         ||||||||||||||||||||||| |||||||| |||||||||||| ||||
Db      81075 CAGATAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAGCAGGAGGTAAAGAAATA 81016

Qy        718 GCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAG 777
              |||||||||||||  |||||||||||||||||| |||||| ||| ||||||||||| |||
Db      81015 GCCAATCATCTATCACCTGAGAGCACAGCGGGAGGGACAATGATCGGGATATAAACCCAG 80956

Qy        778 GCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGT 837
              |||||| |||| | |||||| |||| |||||||||||||||| |||||||||||||||||
Db      80955 GCATTCGAGCCGGTAACAGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT 80896

Qy        838 TTTCACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTA 897
                         ||||||||| |||||| |||||||||||||||||||||| ||||| |||
Db      80895 ----------CTTCACTCTACTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTGTTA 80846

Qy        898 TGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCC 957
               ||||| ||||  ||||| | || || ||||||||||||||||||| ||||   ||||||
Db      80845 CGGCTCGAGCTTCGCTTTCGCTCACCGTCCACCACTGCTGTTTGCCCCCGTGGGAGACCC 80786

Qy        958 GCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGA 1017
               | |||| |||||||||||  ||||||| ||| || ||| ||||||||||||||||||||
Db      80785 ACCGCTGGCTTCCATCCCTCCGGATCCATCAGGGTATCCGCTGTGCTCCTGATCCAGCGA 80726

Qy       1018 GGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGT 1077
              ||  ||||||||||||| ||||| ||| ||| ||| ||||||||||||||| ||||||||
Db      80725 GGCGCCCATTGCCACTCTCGATCGGGCCAAATGCTCGCCATTGTTCCTGCACGGCTAAGT 80666

Qy       1078 GCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCAT 1137
              ||||||||| |||||||| || ||||||||| |||||||||||||| ||||||||||| |
Db      80665 GCCTGGGTTCGTCCTAATCGAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCCGT 80606

Qy       1138 GACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTG 1197
              |||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||
Db      80605 GACCCACGGCTTCTAATAGAGCTATAACACTCACTGCATGGCCCAAGCTTCCATTCCTTG 80546

Qy       1198 GTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              | ||| ||||||||||||||||||||||||||||  |||||||||||||||| | |||||
Db      80545 GAATCCGTGAGGCCAAGAACCCCAGGTCAGAGAACATGAGGCTTGCCACCATCTTGGAAG 80486

Qy       1258 TGGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGAT 1317
              |||||  | | |||||||| |||||||  ||||||||||||||||| |||||||||||
Db      80485 TGGCCTGCGGTCATTTTGGAAGCGGCCTGCCACCATCTTGGGAGCTCTGGGAGCAAGGGC 80426

Qy       1318 CCCCCAGTAACA 1329
              |||||| |||||
Db      80425 CCCCCAATAACA 80414

RESULT 20
US-10-003-806-6/c
; Sequence 6, Application US/10003806
; Publication No. US20020119929A1
; GENERAL INFORMATION:
;  APPLICANT: Bishop, Colin E.
;  APPLICANT:  Agoulnik, Alexander I.
;  APPLICANT:  Zhu, Qichao
;  TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
;  FILE REFERENCE: P02066US1/10024824
;  CURRENT APPLICATION NUMBER: US/10/003,806
;  CURRENT FILING DATE:  2001-11-02
;  PRIOR APPLICATION NUMBER: 60/245,872
;  PRIOR FILING DATE: 2000-11-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 180557
;   TYPE: DNA
;   ORGANISM: Human
US-10-003-806-6

  Query Match             43.6%;  Score 579.8;  DB 13;  Length 180557;
  Best Local Similarity   86.0%;  Pred. No. 2.2e-168;
  Matches  683;  Conservative    0;  Mismatches   90;  Indels   21;  Gaps    3;

Qy        547 ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGG 606
              | ||||||||||||||||||||||||||||||| ||||||||||||| ||||||| ||||
Db      58991 AGTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGG 58932

Qy        607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 666
              |||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||||
Db       58931 GAAGGTGACCGCTTCCACCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 58872

Qy         667 C---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATA 717
               |         ||||||||||||||||||||||| ||||||||||||||||||||| ||||
Db       58871 CAGATAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATA 58812

Qy         718 GCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAG 777
               ||||||||||||||||||||||||||||| ||| |||||| |||  | |||||||| |||
Db       58811 GCCAATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCAGCATATAAACCCAG 58752

Qy         778 GCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGT 837
               ||||   ||||||||||||| || | |||||||||||||||| |||||||||||||||||
Db       58751 GCATCTGAGCCAGCAACAGCTACGCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT 58692

Qy         838 TTTCACTCTATTTCACTCTATTAAATCATGCAACTGCA--CTCTTCTGGTCCGTGTTTTT 895
                          |||||||||||||||| |||| |||||  ||||| |||||    ||| |
Db       58691 ----------CTTCACTCTATTAAATCTTGCAGCTGCACTCTCTTTTGGTCTACATTTGT 58642

Qy         896 TATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGAC 955
                |||| || |||||||||||   ||||| ||||||||||||||||||| | ||| |||||
Db       58641 CATGGTTCGAGCTGAGCTTTCTCTCGCCGTCCACCACTGCTGTTTGCCGCTGTCGCAGAC 58582

Qy         956 CCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGC 1015
               | ||||||||||||||||| |   ||||| |||  |||||  ||||||||||||||||||
Db       58581 CTGCTGCTGACTTCCATCCGTCCAGATCCGGCAAGGTGTCTGCTGTGCTCCTGATCCAGC 58522

Qy        1016 GAGGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAA 1075
               ||||  |||||| |  |||||  || |||||||||||||||| ||||||  || ||||||
Db       58521 GAGGCGCCCATTACAGCTCCCAGTCGGGCTAAAGGCTTGCCACTGTTCCCACACGGCTAA 58462

Qy        1076 GTGCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCC 1135
               ||| | ||||| ||||||||  | ||||||||| |||||||||||||| |||||||||||
Db       58461 GTGTCCGGGTTCGTCCTAATCAAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCC 58402

Qy        1136 ATGACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCT 1195
               |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db       58401 ATGACCCACGGCTTCTAATAGAGCTATAACACTCACTGCATGGCCCAAGATTCCATTCCT 58342

Qy        1196 TGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGA 1255
               ||| ||| ||||||||||||||||||||||||||||   ||||||||||||||| | |||
Db       58341 TGGAATCCGTGAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGA 58282

Qy        1256 AGTGGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGG 1315
               ||  |||| | |||||||||| || ||||||||||||||||||||||| |||||||||||
Db       58281 AGCAGCCCGCCGCCATTTTGGAAGTGGCCCACCACCATCTTGGGAGCTCTGGGAGCAAGG 58222

Qy        1316 ATCCCCCAGTAACA 1329
               | | ||  ||||||
Db       58221 ACCACCTGGTAACA 58208

RESULT 21
US-10-003-806-9/c
; Sequence 9, Application US/10003806
; Publication No. US20020119929A1
; GENERAL INFORMATION:
;  APPLICANT: Bishop, Colin E.
;  APPLICANT:  Agoulnik, Alexander I.
;  APPLICANT:  Zhu, Qichao
;  TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
;  FILE REFERENCE: P02066US1/10024824
;  CURRENT APPLICATION NUMBER: US/10/003,806
;  CURRENT FILING DATE:  2001-11-02
;  PRIOR APPLICATION NUMBER: 60/245,872
;  PRIOR FILING DATE: 2000-11-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
;   LENGTH: 180557
;   TYPE: DNA
;   ORGANISM: Human
US-10-003-806-9
```

```
  Query Match             43.6%;  Score 579.8;  DB 13;  Length 180557;
  Best Local Similarity   86.0%;  Pred. No. 2.2e-168;
  Matches  683;  Conservative    0;  Mismatches   90;  Indels   21;  Gaps    3;

Qy         547 ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGG 606
               | ||||||||||||||||||||||||||||||| ||||||||||||| ||||||| ||||
Db       58991 AGTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGG 58932

Qy         607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 666
               |||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||||
Db       58931 GAAGGTGACCGCTTCCACCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 58872

Qy         667 C---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATA 717
               |         ||||||||||||||||||||||| ||||||||||||||||||||| ||||
Db       58871 CAGATAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATA 58812

Qy         718 GCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAG 777
               ||||||||||||||||||||||||||||| ||| |||||| |||  | |||||||| |||
Db       58811 GCCAATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCAGCATATAAACCCAG 58752

Qy         778 GCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGT 837
               ||||   ||||||||||||| || | |||||||||||||||| |||||||||||||||||
Db       58751 GCATCTGAGCCAGCAACAGCTACGCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT 58692

Qy         838 TTTCACTCTATTTCACTCTATTAAATCATGCAACTGCA--CTCTTCTGGTCCGTGTTTTT 895
                          |||||||||||||||| |||| |||||  ||||| |||||    ||| |
Db       58691 ----------CTTCACTCTATTAAATCTTGCAGCTGCACTCTCTTTTGGTCTACATTTGT 58642

Qy         896 TATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGAC 955
                |||| || |||||||||||   ||||| ||||||||||||||||||| | ||| |||||
Db       58641 CATGGTTCGAGCTGAGCTTTCTCTCGCCGTCCACCACTGCTGTTTGCCGCTGTCGCAGAC 58582

Qy         956 CCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGC 1015
               | ||||||||||||||||| |   ||||| |||  |||||  ||||||||||||||||||
Db       58581 CTGCTGCTGACTTCCATCCGTCCAGATCCGGCAAGGTGTCTGCTGTGCTCCTGATCCAGC 58522

Qy        1016 GAGGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAA 1075
               ||||  |||||| |  |||||  || |||||||||||||||| ||||||  || ||||||
Db       58521 GAGGCGCCCATTACAGCTCCCAGTCGGGCTAAAGGCTTGCCACTGTTCCCACACGGCTAA 58462

Qy        1076 GTGCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCC 1135
               ||| | ||||| ||||||||  | ||||||||| |||||||||||||| |||||||||||
Db       58461 GTGTCCGGGTTCGTCCTAATCAAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCC 58402

Qy        1136 ATGACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCT 1195
               |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db       58401 ATGACCCACGGCTTCTAATAGAGCTATAACACTCACTGCATGGCCCAAGATTCCATTCCT 58342

Qy        1196 TGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGA 1255
               ||| ||| ||||||||||||||||||||||||||||   ||||||||||||||| | |||
Db       58341 TGGAATCCGTGAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGA 58282

Qy        1256 AGTGGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGG 1315
               ||  |||| | |||||||||| || ||||||||||||||||||||||| |||||||||||
Db       58281 AGCAGCCCGCCGCCATTTTGGAAGTGGCCCACCACCATCTTGGGAGCTCTGGGAGCAAGG 58222

Qy        1316 ATCCCCCAGTAACA 1329
               | | ||  ||||||
Db       58221 ACCACCTGGTAACA 58208

RESULT 22
US-10-017-117-1
; Sequence 1, Application US/10017117
; Publication No. US20030124535A1
; GENERAL INFORMATION:
;  APPLICANT: McCarthy, Jeanette
;  TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
;  FILE REFERENCE: MMI-008
;  CURRENT APPLICATION NUMBER: US/10/017,117
;  CURRENT FILING DATE:  2001-12-14
;  PRIOR APPLICATION NUMBER: No. US20030124535A1 yet assigned
;  PRIOR FILING DATE: 2001-12-05
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 161671
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(161671)
;   OTHER INFORMATION: n = a,t,c or g
US-10-017-117-1

  Query Match             42.9%;  Score 570.2;  DB 15;  Length 161671;
  Best Local Similarity   88.5%;  Pred. No. 2.1e-165;
  Matches  655;  Conservative    0;  Mismatches   71;  Indels   14;  Gaps    3;

Qy        547 ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGG 606
              ||||||||||||||||||||||||||||||||| |||||||||||||| ||||||| ||||
Db      74117 ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGG 74176

Qy        607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 666
              ||||||||||||||||| ||||||||| |||||||||||||||||||||||||| ||||||
Db      74177 GAAGGTGACCGCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAACCAAT 74236

Qy        667 CAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCAT 726
              ||||||||||||||||||| |||| |||||||||||||||||||||| ||||||||||||||
Db      74237 CAGAGAGCTCACTAAAATGTTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCAT 74296

Qy        727 CTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAG 786
              |||||||||||||||||||| ||| |||||||||||||||||||||| ||||||||| ||
Db      74297 CTATTGCCTGAGAGCACAGCAGGAGGGACAAGGATTGGGATATAAACCCAGGCATTCGAG 74356

Qy        787 CCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCT 846
              || ||||  |||||||||||||||||||||||| |||||||||||||||||
Db      74357 CCGGCAATGGCAACCCCCTTTGGGTCCCCTCCCCTTGTATGGGAGCTCTGT--------- 74407

Qy        847 ATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAG 906
               ||||||||||||||||| |||||||||||||||||||||| ||||| ||| ||||  ||
Db      74408 -TTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTGTTACGGCTGGAG 74466

Qy        907 CTGAGCTTTTGTTCGCCATCCACCACTGCTG-TTTGCCACCGTCACAGACCCGCTGCTGA 965
              ||||||||| | ||||| ||||||||||||| |||||  ||||  ||||| |||  ||||
Db      74467 CTGAGCTTTCGCTCGCCCTCCACCACTGCTGTTTTGCGGCCGTGGCAGACTCGCCACTGA 74526

Qy        966 CTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCA 1025
              ||||||| | |  | |||| |||| ||||| |||||||||||||||| ||| ||| |||||
Db      74527 CTTCCATTCTTCCGCATCCGGCAGGGTGTCCCCTGTGCTCCTGATCTAGCAAGGCACCCA 74586

Qy       1026 TTGCCACTCCCGATCAGGCTAA---AGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTG 1082
               ||||| ||| |||| ||||||   ||||| |||||||||||||||||||||||||||||
Db      74587 CTGCCATTCCGGATCGGGCTAAATTAGGCTCGCCATTGTTCCTGCATGGCTAAGTGCCTG 74646

Qy       1083 GGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCC 1142
              |||| |||||||| || ||||||||| |||||||||||||||| ||||||||||| ||||||
Db      74647 GGTTCGTCCTAATCGAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCCGTGACCC 74706

Qy       1143 ACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATC 1202
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| |||
Db      74707 ACGGCTTCTAATAGAGCTATAACACTCACCCCATGGCCCAAGATTCCATTCCTTGGAATC 74766

Qy       1203 TGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCC 1262
               ||||||||||||||||||||||||||||   |||||||| |||||| | |||||  | |
Db      74767 CGTGAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGTCACCATCTTGGAAGCAGTC 74826

Qy       1263 CACTGCCATTTTGGTAGCGG 1282
              ||| ||||| |||| ||| |
Db      74827 CACCGCCATCTTGGAAGCTG 74846

RESULT 23
US-10-240-425-1102
; Sequence 1102, Application US/10240425
; Publication No. US20040033502A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Williams, Amanda
;  APPLICANT:  Boland, Joseph F.
;  APPLICANT:  Lord, Reginald V.
;  APPLICANT:  Alvarez, Chris
;  APPLICANT:  Wetzel, Jon C.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Vockley, Joseph G.
;  TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
;  FILE REFERENCE: 44921-5026
;  CURRENT APPLICATION NUMBER: US/10/240,425
;  CURRENT FILING DATE:  2002-09-30
;  PRIOR APPLICATION NUMBER: PCT/US01/09847
;  PRIOR FILING DATE: 2001-03-28
;  PRIOR APPLICATION NUMBER: US 60/193,446
;  PRIOR FILING DATE: 2000-03-31
;  NUMBER OF SEQ ID NOS: 1588
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1102
;   LENGTH: 134292
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20040033502A1 AL031983
US-10-240-425-1102

  Query Match             41.0%;  Score 545.4;  DB 16;  Length 134292;
  Best Local Similarity   85.9%;  Pred. No. 1e-157;
  Matches  680;  Conservative    0;  Mismatches   89;  Indels   23;  Gaps    6;

Qy        547 ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGG 606
              | ||||||||||||||||||||||||||||||| |||||||||||||| ||||||| ||||
Db      54610 AATGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGG 54669

Qy        607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 666
              | ||||||| ||||||| ||||||||| |||||||||||||||||||||||||| |||||||
Db      54670 GGAGGTGACTGCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAAT 54729

Qy        667 C---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATA 717
              |         |||||||||||||||||||||||| |||||||||||||||||||||  |||
Db      54730 CAGGTAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGATATA 54789

Qy        718 GCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAG 777
              |||||||||||||||||||||| |||||||||| |||||| ||| ||||||||||| |||
Db      54790 GCCAATCATCTATTGCCTGAGATCACAGCGGGAGGGACAATGATCGGGATATAAACACAG 54849

Qy        778 GCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGT 837
              |||||| ||||||||||| || |||| |||||||||||||||| |||||||||||||||||
Db      54850 GCATTCGAGCCAGCAAC-GCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT 54908

Qy        838 TTTCACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTA 897
                         ||||||||| |||||| |||||||||||||||||||||||  ||||| |||
Db      54909 ----------CTTCACTCTACTAAATCTTGCAACTGCACTCTTCTGGTCTATGTTTCTTA 54958

Qy        898 TGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCC 957
               ||||| || |||||||| | | |||||||||||||||| |||||||||||||| |||||||
Db      54959 CGGCTCGAGGTGAGCTTTCGCTTGCCATCCACCACTGCCGTTTGCCACCGTCGCAGACCC 55018

Qy        958 GCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGA 1017
              || ||||||||||||||||  |||||  |||| |||||| ||||||| |||| |||| ||
Db      55019 GCGGCTGACTTCCATCCCTC-GGATCTGGCAGGGTGTCCGCTGTGCTTCTGAACCAGTGA 55077

Qy       1018 GGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGT 1077
              ||  ||||||||| |||| |||  |||||||||| | |||||||| |||||| ||||||||
Db      55078 GGCGCCCATTGCCGCTCCTGATTGGGCTAAAGGCGTACCATTGTT-CTGCACGGCTAAGT 55136

Qy       1078 GCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCAT 1137
              |||  ||||  ||||||| || |||||||||| ||||||||| ||||  ||||||||||| |
Db      55137 GCCCAGGTTCTTCCTAATCGAGCTGAACACTAGTCACTGGG-TCCACAGTTCTCTTCCGT 55195

Qy       1138 GACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTG 1197
              |||||| ||||||||||||||||| |||||  |||| |||| |||||||||||||||||||
Db      55196 GACCCATGGCTTCTAATAGAGCTGTAACAACCACCACATGACCCAAGATTCCATTCCTTG 55255
```

```
Qy       1198 GTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
              | |||  |||||||||||||||||||||||||||   ||||||||||||||| | |||||
Db      55256 GAATCCATGAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGAAG 55315

Qy       1258 TGGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGAT 1317
               ||||  | |||||||||| ||| |||||||||||||||| ||||| ||||||||| || 
Db      55316 CGGCCTGCGGCCATTTTGGAAGCAGCCCACCACCATCTTGAGAGCTCTGGGAGCAAAGAC 55375

Qy       1318 CCCCCAGTAACA 1329
              ||||  ||||||
Db      55376 CCCCTGGTAACA 55387


RESULT 24
US-08-979-847-102
; Sequence 102, Application US/08979847
; Publication No. US20030039664A1
;  GENERAL INFORMATION:
;    APPLICANT:  PERRON, HERVE
;    APPLICANT:  BESEME, FREDERIC
;    APPLICANT:  BEDIN, FREDERIC
;    APPLICANT:  PARANHOS-BACCALA, GLAUCIA
;    APPLICANT:  KOMURIAN-PRADEL, FLORENCE
;    APPLICANT:  JOLIVET-REYNAUD, COLETTE
;    APPLICANT:  MANDRAND, BERNARD
;    APPLICANT:  GARSON, JEREMY
;    APPLICANT:  TUKE, PHILIP
;    TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;    TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
;    TITLE OF INVENTION:  THERAPEUTIC PURPOSES
;    NUMBER OF SEQUENCES:  210
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OLIFF & BERRIDGE, PLC
;      STREET:  P.O. BOX 19928
;      CITY:  ALEXANDRIA
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/979,847
;      FILING DATE:  26-NOV-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERRIDGE, WILLIAM P.
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 39046A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6400
;      TELEFAX:  703-836-2787
;  INFORMATION FOR SEQ ID NO:  102:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  635 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-979-847-102

  Query Match             41.0%;  Score 545;  DB 8;  Length 635;
  Best Local Similarity   92.4%;  Pred. No. 7.6e-159;
  Matches  572;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
              ||||||||||| ||| ||||||||||||||||||||||||||||  || ||||||| |||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
              ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436
              ||||||||||||||||| | |||||| ||||||||||||||||||| |||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496
              |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616
              |||||||||||||||||||||||  ||  ||||||||| ||||||| ||||||||||||| 
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676
              ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736
              |||||||||||||| ||||||||| ||||||||||| |||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796
              |||||||||||||| |||||||||| ||||||||||| ||||||||| |||| ||||| |
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856
              ||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy        857 ATTAAATCATGCAACTGCA 875
              |||||||| |||||||| |
Db        601 ATTAAATCTTGCAACTGAA 619


RESULT 25
US-10-114-104-102
; Sequence 102, Application US/10114104
; Publication No. US20030198647A1
;  GENERAL INFORMATION:
;       APPLICANT: PERRON, HERVE
;                  BESEME, FREDERIC
;                  BEDIN, FREDERIC
;                  PARANHOS-BACCALA, GLAUCIA
;                  KOMURIAN-PRADEL, FLORENCE
;                  JOLIVET-REYNAUD, COLETTE
;                  MANDRAND, BERNARD
;                  GARSON, JEREMY
;                  TUKE, PHILIP
;       TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
;                           ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLA
;                           THERAPEUTIC PURPOSES
;       NUMBER OF SEQUENCES: 210
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: OLIFF & BERRIDGE, PLC
;            STREET: P.O. BOX 19928
;            CITY: ALEXANDRIA
;            STATE: VA
;            COUNTRY: USA
;            ZIP: 22320
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/114,104
;            FILING DATE: 03-Apr-2002
;            CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/979,847
;            FILING DATE: 26-NOV-1997
;        ATTORNEY/AGENT INFORMATION:
;            NAME: BERRIDGE, WILLIAM P.
;            REGISTRATION NUMBER: 30,024
;            REFERENCE/DOCKET NUMBER: WPB 39046A
;        TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-836-6400
;            TELEFAX: 703-836-2787
;    INFORMATION FOR SEQ ID NO: 102:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 635 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-10-114-104-102

  Query Match             41.0%;  Score 545;  DB 15;  Length 635;
  Best Local Similarity   92.4%;  Pred. No. 7.6e-159;
  Matches  572;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
              ||||||||||| ||| ||||||||||||||||||||||||||||  || ||||||| |||
Db          1 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTA 60

Qy        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
              ||||| |||||||||||||| | ||||||| ||||||||||||| ||| |||||||||||
Db         61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120

Qy        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436
              ||||||||||||||||| | |||||| ||||||||||||||||||| |||||||||||||
Db        121 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180

Qy        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496
              |||||||||||||||||||||||||||  | |||||||||  ||||||||||||||| ||
Db        181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

Qy        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db        241 CATCAGCCAACCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300

Qy        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616
              |||||||||||||||||||||||  || ||||||||| ||||||| |||||||||||||
Db        301 AGGACTAGCTGGATTTCCTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Qy        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676
              ||||||| || |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Qy        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736
              |||||||||||||| ||||||||| ||||||||||| |||||||||||||||||||||||
Db        421 ACTAAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480

Qy        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796
              |||||||||||||| |||||||||| ||||||||||| ||||||||| |||| ||||| |
Db        481 AGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTCGAGCCGGCAACGG 540

Qy        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTTCACTCTATTTCACTCT 856
              ||||||||||||||||||||||| |||||||| |||||||||||||||||||||||||||
Db        541 CAACCCCCTTTGGGTCCCCTCCCTTTGTATGGCGCTCTGTTTTCACTCTATTTCACTCT 600

Qy        857 ATTAAATCATGCAACTGCA 875
              |||||||| ||||||||| |
Db        601 ATTAAATCTTGCAACTGAA 619

RESULT 26
US-10-220-120-15
; Sequence 15, Application US/10220120
```

```
; Publication No. US20040048253A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  PANZER, Scott R.
;  APPLICANT:  SPIRO, Peter A.
;  APPLICANT:  BANVILLE, Steven C.
;  APPLICANT:  SHAH, Purvi
;  APPLICANT:  CHALUP, Michael S.
;  APPLICANT:  CHANG, Simon C.
;  APPLICANT:  CHEN, Alice
;  APPLICANT:  D'SA, Steven A.
;  APPLICANT:  AMSHEY, Stefan
;  APPLICANT:  DAHL, Christopher R.
;  APPLICANT:  DAM, Tam C.
;  APPLICANT:  DANIELS, Susan E.
;  APPLICANT:  DUFOUR, Gerard E.
;  APPLICANT:  FLORES, Vincent
;  APPLICANT:  FONG, Willy T.
;  APPLICANT:  GREENAWALT, Lila B.
;  APPLICANT:  HILLMAN, Jennifer L.
;  APPLICANT:  JONES, Anissa L.
;  APPLICANT:  LIU, Tommy F.
;  APPLICANT:  ROSEBERRY, Ann M.
;  APPLICANT:  ROSEN, Bruce H.
;  APPLICANT:  RUSSO, Frank D.
;  APPLICANT:  STOCKDREHER, Theresa K.
;  APPLICANT:  DAFFO, Abel
;  APPLICANT:  WRIGHT, Rachel J.
;  APPLICANT:  YAP, Pierre E.
;  APPLICANT:  YU, Jimmy Y.
;  APPLICANT:  BRADLEY, Diana L.
;  APPLICANT:  BRATCHER, Shawn R.
;  APPLICANT:  CHEN, Wensheng
;  APPLICANT:  COHEN, Howard J.
;  APPLICANT:  HODGSON, David M.
;  APPLICANT:  LINCOLN, Stephen E.
;  APPLICANT:  JACKSON, Stuart
;  TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PT-1113 PCT
;  CURRENT APPLICATION NUMBER: US/10/220,120
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774;
;    60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;
;    60/184,769; 60/184,768; 60/184,837; 60/184,697; 60/184,841;
;    60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;
;    60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;
;    60/204,525; 60/205,285; 60/205,232; 60/205,323; 60/205,287;
;    60/205,324; 60/205,286
;  PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
;    2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-05-17;
;    2000-05-17; 2000-05-12; 2000-05-16; 2000-05-16; 2000-05-15;
;    2000-05-16; 2000-05-17; 2000-05-16; 2000-05-17; 2000-05-17;
;    2000-05-17; 2000-05-17
;  NUMBER OF SEQ ID NOS: 422
;  SOFTWARE: PERL Program
; SEQ ID NO 15
;   LENGTH: 849
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20040048253A1 LG:429446.1:2000FEB18
US-10-220-120-15

  Query Match             40.3%;  Score 536.2;  DB 16;  Length 849;
  Best Local Similarity   81.4%;  Pred. No. 5e-156;
  Matches  650;  Conservative    0;  Mismatches  135;  Indels   14;  Gaps    2;

Qy        208 CCCCTTCTTAGGACCTCTAGCAGCTATAATATTTTTACTCCTCTTTGGACCCTGTATCTT 267
              | || |||| ||| |||||  | | |  |           |  || |   |   | |
Db         46 CACCATCTTGGGAGCTCTAAGAACAAAGACCCACCCGTAACATTTGGTGGCAACTGTACG 105
```

```
Qy        268 CAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAAATAGTTCT 327
                  |||        ||||||||||||||||||  || ||||||| |||||||| |||||
Db        106 GGGATTCTCCAAAGTGTTTGTCTCTTCCAGAATCAAAACTGTAAAACTACAAATTGTTCT 165

Qy        328 TCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCCCTGGACCGGCC 387
              ||||||||| |||||||||||||| |||||||| |||||||   ||||| ||||||||||
Db        166 TCAAATGGAGCCCCAGATGCAGTCTATGACTAAGATCTACCACAGACCCTTGGACCGGCC 225

Qy        388 TGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAGGAAATCTCAAC 447
              |||||| | |||||| |||||| ||||||| |||| | |||||||| |||||||||||||
Db        226 TGCTAGCCCATGCTCCGATGTTGATGACATCGAAGGCGCCCCTCCCAAGGAAATCTCAAC 285

Qy        448 TGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGTTGTCAGCCAAC 507
              ||||  ||||||||||| | ||||||| | |||||||||||||||||||  |  ||||||
Db        286 TGCATGACCCCTACTACGCCCCAATTCCGCAGGAAGCAGTTAGAGCAGTCATTGGCCAAC 345

Qy        508 CTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGACAGGACTAGCTG 567
              ||||||||||| ||||| |||||||||||||||||| | |||||||||||||||||||||
Db        346 CTCCCCAACAGCACTTGAGTTTTCCTGTTGAGAGGGAGAACTGAGAGACAGGACTAGCTG 405

Qy        568 GATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACCGCATCCATCTT 627
              ||||||| ||||  |||||||||||| ||||||| ||||||||||||||| ||||| |||
Db        406 GATTTCCCAGGCCAACTAAGAATCCCTAAGCCTATCTGGGAAGGTGACCGTATCCACCTT 465

Qy        628 TAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCA---------GAGAGCTCAC 678
              ||||||| ||||||||||||||||||||||||||| |||||         ||||||||||
Db        466 TAAACATAGGGCTTGCAACTTAGCTCACACCCGACTAATCAGATAGTAAGGAGAGCTCAC 525

Qy        679 TAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTGAG 738
              |||||||||||| ||||||||||||||||||||| |||||||||||||||||||||||||
Db        526 TAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAG 585

Qy        739 AGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAGCA 798
              |||||||| ||| |||||| |||  |||||||||| |||||||||||||||||||  || 
Db        586 AGCACAGCAGGAGGGACAATGATCAGGATATAAACCCAGGCATTCAAGCCAGCAATGGCT 645

Qy        799 ACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCTAT 858
              |||| |||||||||||||||  ||||||||||||||||||||||  |||||| |   | |
Db        646 ACCCTCTTTGGGTCCCCTCCGTTTGTATGGGAGCTCTGTTTTCATGCTATTTTAAAGTCT 705

Qy        859 TAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCTGAGCTTTTGT 918
              |  | |         |||||||||| ||| ||||| |||  |||  ||||||||||||| 
Db        706 TGCAAC-----TGCACACTCTTCTGATCCATGTTTGTTACAGCTTGAGCTGAGCTTTTGC 760

Qy        919 TCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTTCCATCCCTTT 978
              || ||||||| || |||| ||||||||| ||||||||||||  ||||||||||||||| |
Db        761 TCACCATCCAGCAATGCTCTTTGCCACCATCACAGACCCGCCACTGACTTCCATCCCTCT 820

Qy        979 GGATCCAGCAGAGTGTCCA 997
              ||||||||||| |||| ||
Db        821 GGATCCAGCAGGGTGTTCA 839


RESULT 27
US-10-719-993-6792/c
; Sequence 6792, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6792
;   LENGTH: 366710
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(366710)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6792

  Query Match             39.9%;  Score 530.6;  DB 18;  Length 366710;
  Best Local Similarity   91.0%;  Pred. No. 7.3e-153;
  Matches  563;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
              ||||||||||| |||  ||||||||||||||| |||||||||||| || ||||||| |||
Db      43899 CCCTGTATCTTTAACCACCTTGTTAAGTTTGTGTCTTCCAGAATTAAAACTGTAAAACTA 43840

Qy        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
              ||||| ||||||||||||||||||||||||||||||||| |||| |||| ||  ||||||
Db      43839 CAAATCGTTCTTCAAATGGAACCCCAGATGCAGTCCATGTCTAAGATCTCCCACGGACCC 43780

Qy        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436
              || ||| |||||||||| | |||||||| |||||||||||| |||||||| ||||| |||
Db      43779 CTAGACTGGCCTGCTAGCCCATGCTCTGGTGTTAATGACATCGAAGTCACTCCTCCTGAG 43720

Qy        437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496
              |||||||||||||||||||||||||||  | ||| ||||| |||||||||||||||  ||
Db      43719 GAAATCTCAACTGCACAACCCCTACTATGCCCCAGTTCAGCAGGAAGCAGTTAGAGTGGT 43660

Qy        497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556
                |||||||||||| ||||||| |||||||||||||||||||||| | ||||||||||||
Db      43659 CATCAGCCAACCTCACCAACAGCACTTGGGTTTTCCTGTTGAGAGAGGGGACTGAGAGAC 43600

Qy        557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616
              ||||||||||||||||||||||| || |||||||||| || |||| ||||||||||||||
Db      43599 AGGACTAGCTGGATTTCCTAGGCCGATTAAGAATCCCTAAACCTAGCTGGGAAGGTGACC 43540

Qy        617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676
              ||||||| ||||||||| ||||||||||||||||||||||||| ||||||||||||||||
Db      43539 GCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCAACCAATCAGAGAGCTC 43480

Qy        677 ACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTG 736
              |||||||||||||| ||||||||||||||||||||| |||||||||||||||||||||||
Db      43479 ACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 43420

Qy        737 AGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAG 796
              ||||||||| |||| ||||||||||||  |||||||| ||||||||| |||||||||| |
Db      43419 AGAGCACAGTGGGAGGGACAAGGATTGCAATATAAACCCAGGCATTCGAGCCAGCAACGG 43360

Qy        797 CAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 856
              ||||| |||||||||||| |||| ||||||||||||||||||||||||||||||||||||
Db      43359 CAACCGCCTTTGGGTCCCTTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCT 43300

Qy        857 ATTAAATCATGCAACTGCA 875
              |||||||| |||||||| |
Db      43299 ATTAAATCTTGCAACTGAA 43281


RESULT 28
US-10-264-237-2833/c
; Sequence 2833, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA131P1
;  CURRENT APPLICATION NUMBER: US/10/264,237
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/16450
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,515
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 2876
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2833
;   LENGTH: 17758
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
US-10-264-237-2833

  Query Match             38.5%;  Score 512;  DB 16;  Length 17758;
  Best Local Similarity   84.7%;  Pred. No. 9.1e-148;
  Matches  614;  Conservative    0;  Mismatches   93;  Indels   18;  Gaps    3;

Qy         547 ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGG 606
               | |||||||||||||||||||||| ||| |||| |||||| |||||| ||||||| ||||
Db        6498 AGTGAGAGACAGGACTAGCTGGATGTCCGAGGCCGACTAATAATCCCTAAGCCTAGCTGG 6439

Qy         607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 666
               ||||||||||||||||| ||||||||| ||||| ||||| ||||||||||||| ||||||
Db        6438 GAAGGTGACCGCATCCACCTTTAAACACGGGGCATGCAATTTAGCTCACACCCAACCAAT 6379

Qy         667 C---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATA 717
               |         ||||||||||||||| ||||| | ||||||||| |||| |||||| ||||
Db        6378 CAGGTAGTAAAGAGAGCTCACTAAAGTGCTAGTTAGGCAAAAATAGGAAGTAAAGAAATA 6319

Qy         718 GCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAG 777
               |||||||||||||  ||||||||||||| || | |||||| |||  ||||| |||| |||
Db        6318 GCCAATCATCTATCACCTGAGAGCACAGGGGAAGGGACAATGATCTGGATAGAAACCCAG 6259

Qy         778 GCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGT 837
               |||||| |||||||||| || |||| |||||||||||||||| |||||||||| ||||||
Db        6258 GCATTCCAGCCAGCAACGGCTACCCACTTTGGGTCCCCTCCCGTTGTATGGGAACTCTGT 6199

Qy         838 TTTCACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTA 897
               ||||||||||||  |   |   |   ||        |||||||||||||||||| | |||
Db        6198 TTTCACTCTATTAAAACTTGAAACTGCA--------CACTCTTCTGGTCCGTGTCTGTTA 6147

Qy         898 TGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCC 957
                ||||  ||||||||||| | ||||| ||||||||||||||||||| ||||| |||||||
Db        6146 CGGCTTGAGCTGAGCTTTCGCTCGCCGTCCACCACTGCTGTTTGCCGCCGTCGCAGACCC 6087

Qy         958 GCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGA 1017
               || ||||||||||| ||||  |||||| |||| ||||||||||||||| |||||||| ||
Db        6086 GCCGCTGACTTCCA-CCCTCCGGATCCGGCAGGGTGTCCACTGTGCTCGTGATCCAGGGA 6028

Qy        1018 GGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGT 1077
               |  |||||||||| ||||| ||| ||||| ||||| |||||||||||||||  ||||| |
Db        6027 GACACCCATTGCCGCTCCCAATCGGGCTAGAGGCTCGCCATTGTTCCTGCACAGCTAACT 5968

Qy        1078 GCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCAT 1137
               |||||||||  |||| || || ||||||||||||| |||||||||| ||||||||||| |
Db        5967 GCCTGGGTTCATCCTGATCGAGCTGAACACTGGTCGCTGGGTTCCACGGTTCTCTTCCGT 5908

Qy        1138 GACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTG 1197
               |||||||||||||||||||||||| ||||||||||||| |||||| ||||||||||||||
Db        5907 GACCCACGGCTTCTAATAGAGCTAAAACACTCACCGCAGGGCCCAGGATTCCATTCCTTG 5848

Qy        1198 GTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAG 1257
               | ||| ||||||||||||||||||||||||||||   |||||||||  || | | |||||
Db        5847 GAATCCGTGAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCTGCCCTCTTGGAAG 5788

Qy        1258 TGGCC 1262
               |||||
Db        5787 TGGCC 5783

RESULT 29
US-09-731-231A-3/c
; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1
; GENERAL INFORMATION:
;  APPLICANT: GUEGLER, Karl et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001007
;  CURRENT APPLICATION NUMBER: US/09/731,231A
;  CURRENT FILING DATE:  2000-12-07
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 3
;   LENGTH: 326014
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(326014)
;   OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3

  Query Match             37.8%;  Score 502;  DB 9;  Length 326014;
  Best Local Similarity   88.7%;  Pred. No. 5.7e-144;
  Matches  557;  Conservative    0;  Mismatches   62;  Indels    9;  Gaps    1;

Qy         257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
               ||||||||||| ||| |||||||||||||||||||||||| |||||||||| |||| |||
Db      170822 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAAAATTGAAGCTCTAAAACTA 170763

Qy         317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
               ||||| |||||||||||||| ||||||| ||||||||||||||| ||||||||  |||||
Db      170762 CAAATTGTTCTTCAAATGGAGCCCCAGACGCAGTCCATGACTAAGATCTACCGCAGACCC 170703

Qy         377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436
               ||||||||||||||||| | |||||| |||||||||||||| |||| ||||||||| |||
Db      170702 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCTCCAGAG 170643

Qy         437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496
               ||||||||||||||| |||||| ||||| | ||||||||| |||||||||||||||||||
Db      170642 GAAATCTCAACTGCATAACCCCCACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 170583

Qy         497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556
                 || |||| |||||||||||| |||||||||||||||||||||||| | ||| ||||||
Db      170582 CATCGGCCATCCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTCAGAGAC 170523

Qy         557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616
               ||||||||||||||||||||||||||||||||||||| ||||||| |||||||||||||
Db      170522 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 170463

Qy         617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 667
               || || | ||||||||  |||||||||||||||||||||||| ||||||||
Db      170462 GCTTCTACCTTTAAACCCGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGGTAGGAA 170403

Qy         668 AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATC 727
               |||||||||||||||||| |||  |||| |||||||||||||||| ||||||||||||||
Db      170402 AGAGAGCTCACTAAAATGGTAACTAGGCTAAAACAGGAGGTAAAGAAATAGCCAATCATC 170343

Qy         728 TATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGC 787
               ||| |||||||| ||||| |||| |||||| ||||||| ||||||| ||||||||  |||
Db      170342 TATCGCCTGAGAACACAGTGGGAGGGACAATGATTGGGTTATAAACCCAGGCATTGGAGC 170283

Qy         788 CAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTA 847
               ||||||  || |||   ||||||||||||||| |||||||||||||||||||||||||||
Db      170282 CAGCAATGGCTACCATTTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTA 170223

Qy         848 TTTCACTCTATTAAATCATGCAACTGCA 875
               ||||||||||||||||| |||||| |||
Db      170222 TTTCACTCTATTAAATCTTGCAACAGCA 170195

RESULT 30
US-10-751-985-3/c
; Sequence 3, Application US/10751985
; Publication No. US20040126861A1
; GENERAL INFORMATION:
;  APPLICANT: GUEGLER, Karl et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001007CON
;  CURRENT APPLICATION NUMBER: US/10/751,985
;  CURRENT FILING DATE:  2004-01-07
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 326014
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(326014)
;   OTHER INFORMATION: n = A,T,C or G
US-10-751-985-3

  Query Match             37.8%;  Score 502;  DB 17;  Length 326014;
  Best Local Similarity   88.7%;  Pred. No. 5.7e-144;
  Matches  557;  Conservative    0;  Mismatches   62;  Indels    9;  Gaps    1;

Qy         257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
               ||||||||||| ||| ||||||||||||||||||||||||| |||||||||| |||| |||
Db      170822 CCCTGTATCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAAAATTGAAGCTCTAAAACTA 170763

Qy         317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
               ||||| |||||||||||||| ||||||| ||||||||||||||| ||||||||  |||||
Db      170762 CAAATTGTTCTTCAAATGGAGCCCCAGACGCAGTCCATGACTAAGATCTACCGCAGACCC 170703

Qy         377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436
               ||||||||||||||||| | |||||| |||||||||||||| |||| ||||||||| |||
Db      170702 CTGGACCGGCCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCTCCAGAG 170643

Qy         437 GAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496
               ||||||||||||||| |||||| ||||| | ||||||||| |||||||||||||||||||
Db      170642 GAAATCTCAACTGCATAACCCCCACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCAGT 170583

Qy         497 TGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGAC 556
                 || |||| |||||||||||| |||||||||||||||||||||||| | ||| ||||||
Db      170582 CATCGGCCATCCTCCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTCAGAGAC 170523

Qy         557 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACC 616
               ||||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||| 
Db      170522 AGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT 170463

Qy         617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC--------- 667
               || || | ||||||||  |||||||||||||||||||||||| ||||||||
Db      170462 GCTTCTACCTTTAAACCCGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGGTAGGAA 170403

Qy         668 AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATC 727
               |||||||||||||||||| |||  |||| |||||||||||||||| ||||||||||||||
Db      170402 AGAGAGCTCACTAAAATGGTAACTAGGCTAAAACAGGAGGTAAAGAAATAGCCAATCATC 170343

Qy         728 TATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGC 787
               ||| |||||||| ||||| |||| |||||| ||||||| ||||||| ||||||||  |||
Db      170342 TATCGCCTGAGAACACAGTGGGAGGGACAATGATTGGGTTATAAACCCAGGCATTGGAGC 170283

Qy         788 CAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTA 847
               ||||||  || |||   ||||||||||||||| |||||||||||||||||||||||||||
Db      170282 CAGCAATGGCTACCATTTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTA 170223

Qy         848 TTTCACTCTATTAAATCATGCAACTGCA 875
               ||||||||||||||||| |||||| |||
Db      170222 TTTCACTCTATTAAATCTTGCAACAGCA 170195

Search completed: January 22, 2005, 20:08:44
Job time : 731.788 secs
```

```
                       GenCore version 5.1.6
              Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 22, 2005, 12:01:19 ; Search time 3958.84 Seconds
                                             (without alignments)
                                             12232.983 Million cell updates/sec

Title:          US-09-319-156B-12
Perfect score:  1329
Sequence:       1 tcaaaatcgaagagctttag..........gcaaggatcccccagtaaca 1329

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters:      65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_htc:*
                4:  gb_est3:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_gss1:*
                9:  gb_gss2:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                    %
Result             Query
   No.    Score    Match Length DB  ID                    Description
 ---------------------------------------------------------------------------
     1    706.8    53.2    2748  3  CR605851              CR605851 full-leng
     2    702.8    52.9    2749  3  CR617248              CR617248 full-leng
     3      702    52.8    2500  3  CR622175              CR622175 full-leng
     4      699    52.6    2748  3  CR625046              CR625046 full-leng
c    5    684.2    51.5    1071  5  BX365066              BX365066 BX365066
     6    678.6    51.1    2716  3  CR613169              CR613169 full-leng
c    7    669.4    50.4     998  5  BX337769              BX337769 BX337769
c    8    661.2    49.8    1058  5  BX378303              BX378303 BX378303
c    9    652.2    49.1    1019  5  BX439636              BX439636 BX439636
c   10      610    45.9     966  5  BX380176              BX380176 BX380176
c   11    608.2    45.8     758  5  BX357208              BX357208 BX357208
    12    593.4    44.7     771  4  BI087886              BI087886 602852690
    13      592    44.5    4204  3  BC035153              BC035153 Homo sapi
    14      561    42.2     719  7  CN272394              CN272394 170006000
c   15      544    40.9     900  5  BX459153              BX459153 BX459153
c   16    500.6    37.7    1002  5  BX953822              BX953822 DKFZp781M
c   17    480.8    36.2     679  9  AG113694              AG113694 Pan trogl
c   18    480.8    36.2     999  4  BM918330              BM918330 AGENCOURT
    19    475.2    35.8     683  9  AG134524              AG134524 Pan trogl
    20      474    35.7     669  9  AG049953              AG049953 Pan trogl
    21      473    35.6     681  9  AG116971              AG116971 Pan trogl
    22    469.6    35.3     710  9  AG096341              AG096341 Pan trogl
    23    468.2    35.2    1160  3  BC030968              BC030968 Homo sapi
    24    465.2    35.0     893  5  BQ437925              BQ437925 AGENCOURT
```

```
    25    456.2    34.3     670  9  AG072852              AG072852 Pan trogl
    26      455    34.2     712  9  AG099717              AG099717 Pan trogl
c   27    439.2    33.0     689  9  AG121669              AG121669 Pan trogl
c   28    438.8    33.0     921  8  BH149565              BH149565 ENTQG48TR
c   29    435.6    32.8     679  9  AG076758              AG076758 Pan trogl
    30    433.6    32.6     683  9  AG102951              AG102951 Pan trogl
    31    430.6    32.4     611  8  AQ381711              AQ381711 RPCI11-16
c   32    429.4    32.3     736  9  AG086046              AG086046 Pan trogl
    33    427.6    32.2     714  8  AQ238712              AQ238712 RPCI11-64
    34    424.4    31.9     697  9  AG108186              AG108186 Pan trogl
    35    423.8    31.9     675  9  AG117095              AG117095 Pan trogl
    36    422.2    31.8     797  6  CB990962              CB990962 AGENCOURT
    37    422.2    31.8     803  6  CB992422              CB992422 AGENCOURT
    38    422.2    31.8     826  6  CB990802              CB990802 AGENCOURT
    39    422.2    31.8     837  6  CB996320              CB996320 AGENCOURT
    40    422.2    31.8     852  6  CB994833              CB994833 AGENCOURT
c   41    419.4    31.6     672  9  AG040241              AG040241 Pan trogl
c   42    418.8    31.5     611  7  CR542951              CR542951 DKFZp470A
    43    418.2    31.5     870  7  CN645411              CN645411 ILLUMIGEN
c   44    416.2    31.3     692  9  AG114407              AG114407 Pan trogl
c   45    415.6    31.3     494  1  AA781423              AA781423 aj26c03.s

                                   ALIGNMENTS


RESULT 1
CR605851
LOCUS       CR605851                2748 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DE012YJ24 of Placenta of Homo sapiens
            (human).
ACCESSION   CR605851
VERSION     CR605851.1  GI:50486658
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2748)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2748)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2748
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YJ24"
                     /tissue_type="Placenta"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             53.2%;  Score 706.8;  DB 3;  Length 2748;
  Best Local Similarity   90.5%;  Pred. No. 1.1e-202;
  Matches  769;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db       1911 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1970
```

```
Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        1971 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2030

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        2031 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2090

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        2091 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2150

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        2151 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2210

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        2211 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2258

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        2259 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2318

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        2319 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2378

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        2379 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2438

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        2439 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2498

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2499 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2558

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        2559 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2618

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        2619 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2678

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        2679 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2738

Qy         841 CACTCTATTT 850
               ||  ||||||
Db        2739 CATGCTATTT 2748


RESULT 2
CR617248
LOCUS       CR617248                2749 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI022YJ18 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR617248
VERSION     CR617248.1  GI:50498055
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2749)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2749)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .2749
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI022YJ18"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             52.9%;  Score 702.8;  DB 3;  Length 2749;
  Best Local Similarity   90.4%;  Pred. No. 1.8e-201;
  Matches  765;  Conservative    0;  Mismatches   69;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        1916 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1975

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        1976 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2035

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        2036 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2095

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        2096 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2155

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        2156 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2215

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        2216 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2263

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        2264 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2323

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        2324 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2383

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        2384 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2443

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        2444 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2503

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2504 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2563
```

```
Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        2564 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2623

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        2624 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2683

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        2684 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2743

Qy         841 CACTCT 846
               ||  ||
Db        2744 CATGCT 2749


RESULT 3
CR622175
LOCUS       CR622175                2500 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI051YM13 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR622175
VERSION     CR622175.1  GI:50502982
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2500)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2500)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .2500
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             52.8%;  Score 702;  DB 3;  Length 2500;
  Best Local Similarity   90.6%;  Pred. No. 3.1e-201;
  Matches  763;  Conservative    0;  Mismatches   67;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        1669 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1728

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        1729 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1788

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        1789 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 1848
```

```
Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        1849 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1908

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        1909 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 1968

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        1969 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2016

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        2017 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2076

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        2077 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2136

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        2137 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2196

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        2197 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2256

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2257 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2316

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        2317 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2376

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        2377 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2436

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||| ||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db        2437 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTT 2496

Qy         841 CA 842
               ||
Db        2497 CA 2498


RESULT 4
CR625046
LOCUS       CR625046                2748 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI044YK06 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR625046
VERSION     CR625046.1  GI:50505853
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2748)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2748)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .2748
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI044YK06"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             52.6%;  Score 699;  DB 3;  Length 2748;
  Best Local Similarity   90.6%;  Pred. No. 2.6e-200;
  Matches  760;  Conservative    0;  Mismatches   67;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        1922 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1981

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        1982 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2041

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        2042 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2101

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        2102 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2161

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        2162 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2221

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        2222 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2269

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        2270 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2329

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        2330 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2389

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        2390 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2449

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        2450 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2509

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2510 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2569

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        2570 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2629

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        2630 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2689

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTT 839
               ||| |||| ||||| |||||||||||||||||||||||| |||||||||||||||||||
Db        2690 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGTATGGAGCTCTGTTT 2748


RESULT 5
BX365066/c
LOCUS       BX365066                1071 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX365066 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI046YA18 3-PRIME, mRNA sequence.
ACCESSION   BX365066
VERSION     BX365066.2  GI:46304105
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1071)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30374869.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS1AI012ZC10NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1. .1071
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI046YA18"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             51.5%;  Score 684.2;  DB 5;  Length 1071;
  Best Local Similarity   89.9%;  Pred. No. 6.1e-196;
  Matches  755;  Conservative    3;  Mismatches   69;  Indels   13;  Gaps    2;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db         827 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 768

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db         767 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 708

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db         707 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 648

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         647 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 588

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db         587 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 528

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db         527 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 480

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                |||||||| |||||||||||||||||||||||| | | | |||||||||||||||||  |
Db         479 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 420

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db         419 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 360

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db         359 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 300

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db         299 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 240

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db         239 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 180

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db         179 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 120

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |  
Db         119 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 60

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
               ||| |||| ||||| ||||||||||||||  |||||||| ||||||| ||::||:|||||
Db          59 TTCGAGCCGGCAACGGCAACCCCCTTTGGTNCCCCTCCC-TTGTATGCGASMTCKGTTTT 1

RESULT 6
CR613169
LOCUS       CR613169                2716 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DE013YI20 of Placenta of Homo sapiens
            (human).
ACCESSION   CR613169
VERSION     CR613169.1  GI:50493976
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2716)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2716)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..2716
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE013YI20"
                     /tissue_type="Placenta"
                     /plasmid="pCMVSPORT_6"
```

```
ORIGIN

  Query Match             51.1%;  Score 678.6;  DB 3;  Length 2716;
  Best Local Similarity   90.5%;  Pred. No. 4.1e-194;
  Matches  739;  Conservative    0;  Mismatches   66;  Indels   12;  Gaps    1;

Qy           1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
               |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        1912 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 1971

Qy          61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
               |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        1972 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2031

Qy         121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
               |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        2032 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 2091

Qy         181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        2092 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 2151

Qy         241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                 ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        2152 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 2211

Qy         301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        2212 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2259

Qy         361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                |||||||| |||||||||||||||||||||||| | | | |||||||||||||||||  |
Db        2260 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 2319

Qy         421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
               || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        2320 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 2379

Qy         481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
               ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        2380 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 2439

Qy         541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                || ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        2440 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCT 2499

Qy         601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
               | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        2500 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2559

Qy         661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               |||||||||||||||||||||||||||||| ||||||| ||||||||||||| |||||||
Db        2560 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCC 2619

Qy         721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
               |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |  
Db        2620 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 2679

Qy         781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTC 817
               ||| |||| ||||| ||||||||||||||||||||||
Db        2680 TTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTC 2716


RESULT 7
BX337769/c
LOCUS       BX337769                 998 bp    mRNA    linear   EST 07-APR-2004
DEFINITION  BX337769 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YM13 3-PRIME, mRNA sequence.
ACCESSION   BX337769
VERSION     BX337769.2  GI:46272079
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 998)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 2, 2003 this sequence version replaced gi:30337641.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DI051AG07NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..998
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI051YM13"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             50.4%;  Score 669.4;  DB 5;  Length 998;
  Best Local Similarity   89.2%;  Pred. No. 1.9e-191;
  Matches  754;  Conservative    2;  Mismatches   74;  Indels   15;  Gaps    3;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        835 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 776

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        775 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 716

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        715 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 656

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        655 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 596

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        595 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 536

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        535 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 488

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  ||||||||||||||||||||:| | | | |||||||||||||||||  |
Db        487 GATCTACCGCAGACCCCTGGACCGGCCTGCTRGCCCACGATCTGATGTTAATGACATCAA 428

Qy        421 A-GTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAG 479
              | | ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| ||
Db        427 ACGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAG 368

Qy        480 GAAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAG 539
              |||||||||||||| || ||| ||||||||||||||||| |||| |||||||||||||||
Db        367 GAAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAG 308

Qy        540 AGGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCC 599
              | || ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        307 ATGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCC 248

Qy        600 TANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCC 659
              || |||||||||||||| |||||| ||||||||| ||||||||||||||:|||||||||
Db        247 TAGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTRGCTCACACCT 188

Qy        660 GACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGC 719
              ||||||||||||||||||||||||||||||| ||||||| ||||||||||||| ||||||
Db        187 GACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGC 128

Qy        720 CAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGC 779
              ||||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        127 CAATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGT 68

Qy        780 ATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATT--GTATGGGAGCTCTGT 837
               ||| |||| ||||| |||||||||||||||||||  | | ||     |||||| |||||
Db         67 CTTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCNNCNCCTTTGTANTGGGAGGTCTGT 8

Qy        838 TTTCA 842
              |||||
Db          7 TTTCA 3


RESULT 8
BX378303/c
LOCUS       BX378303                1058 bp    mRNA    linear   EST 23-APR-2004
DEFINITION  BX378303 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI018YH07 3-PRIME, mRNA sequence.
ACCESSION   BX378303
VERSION     BX378303.2  GI:46557492
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1058)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 8, 2003 this sequence version replaced gi:30439129.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DI018CD04NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..1058
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI018YH07"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             49.8%;  Score 661.2;  DB 5;  Length 1058;
  Best Local Similarity   86.5%;  Pred. No. 6e-189;
  Matches  727;  Conservative    9;  Mismatches   92;  Indels   12;  Gaps    1;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
```

```
              |||||||||||||||||||||||||||||||||  |||||||:::|:|:| ||| |||||
Db        832 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGKKKAWCYTTTTTTTTTTT 773

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| ::|:|| | |||||||||||||||||
Db        772 AGGGGAAGAATGCTGTTACTATGTTAATCAATCCKKAWTCKTCACTGAGAAAGTTAAAGA 713

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        712 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 653

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        652 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 593

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        592 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 533

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        532 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 485

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        484 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 425

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||||||||| ||||||| | ||||||||| |||
Db        424 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGG 365

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| ||||||||||||||||| |||| ||||||||||||||||
Db        364 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGA 305

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || |||||||||||||||||||||||||||||:||||||||||||||||||| ||||||
Db        304 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCYTAGGCTGACTAAGAATCCCTAAGCCT 245

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db        244 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 185

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
              |||||||||||||||||||||||||||||  ||||||| ||||||||||||| |||||||
Db        184 ACCAATCAGAGAGCTCACTAAAATGCTAACTAGGCAAAGACAGGAGGTAAAGAAATAGCC 125

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              |||||||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db        124 AATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTC 65

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
              ||| |||| ||||| |||||||  |   |||  |       | | || |   | |  |||
Db         64 TTCGAGCCGGCAACGGCAACCCNNTNNTGGTNNCNNTGGCGTTTGTGTGGAGTTTTGTTT 5


RESULT 9
BX439636/c
LOCUS       BX439636                1019 bp    mRNA    linear   EST 04-MAY-2004
DEFINITION  BX439636 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012YJ24
            3-PRIME, mRNA sequence.
ACCESSION   BX439636
VERSION     BX439636.2  GI:47000005
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1019)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 15, 2003 this sequence version replaced gi:30771765.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DE012DE12NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1. .1019
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YJ24"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             49.1%;  Score 652.2;  DB 5;  Length 1019;
  Best Local Similarity   85.5%;  Pred. No. 3.2e-186;
  Matches  725;  Conservative   19;  Mismatches   91;  Indels   13;  Gaps    2;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        835 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 776

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db        775 AGGGGAAGAATGCTGTTACTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 716

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        715 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 656

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        655 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 596

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db        595 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 536

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| |||            |||||||| ||| |||||||||||| ||||||
Db        535 CGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAA 488

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               ||||||||  |||||||||||||||||||||| | | | |||||||||||||||||  |
Db        487 GATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAA 428

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              || ||||||||| |||||||||||| |||:||||||  |||||| | ||||||||| |||
Db        427 AGGCACCCCTCCTGAGGAAATCTCAGCTGMACAACCTATACTACGCCCCAATTCAGCAGG 368

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || ||| |||||||||||| |||| |||| ||||||||||||||||
Db        367 AAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCCACAGCACTTAGGTTTTCCTGTTGAGA 308

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               || ||||||||||||||||||||||||||: ||:|||||||||||||||||| ||||||
Db        307 TGGGGGACTGAGAGACAGGACTAGCTGGATYCCCYAGGCTGACTAAGAATCCCTAAGCCT 248

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              | |||||||||||||| |||||| ||||||||| |||||||||||||||||||||||| |
Db            247 AGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 188

Qy            661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
                  ||||||||||||||| |||||||||||||  ||||||| |||||||||:::: ||||: :
Db            187 ACCAATCAGAGAGCTKACTAAAATGCTAACTAGGCAAAKACAGGAGGTWWWKTAATAKNV 128

Qy            721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
                  :|| || ||||||||||||||||||| ||| ||||:| :|| ||||||||||: || |
Db            127 RATTATTTATTGCCTGAGAGCACAGCAGGAGGGACWATKATCGGGATATAAAYCCAAGTC 68

Qy            781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTT 840
                  ||| |||| ||||||| ||||:::|||   ||:|: |   |||| |||| : | | |||
Db             67 TTCGAGCCGGCAACAGAAACCYYYTTT-TTTCYCYCCTTTTTGTNTGGGNDTTTTTTTTN 9

Qy            841 CACTCTAT 848
                     | | |
Db              8 NTTTATTT 1


RESULT 10
BX380176/c
LOCUS       BX380176                 966 bp    mRNA    linear   EST 28-APR-2004
DEFINITION  BX380176 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI044YK06 3-PRIME, mRNA sequence.
ACCESSION   BX380176
VERSION     BX380176.2  GI:46834414
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 966)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 8, 2003 this sequence version replaced gi:30460243.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DI044BF03NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1..966
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI044YK06"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             45.9%;  Score 610;  DB 5;  Length 966;
  Best Local Similarity   82.5%;  Pred. No. 2.1e-173;
  Matches  693;  Conservative   24;  Mismatches  110;  Indels   13;  Gaps    2;

Qy              1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
                  |||||||||||||||||||:||||||||:||||  |||||||||||||||||||||||||
Db            830 TCAAAATCGAAGAGCTTTARACTTGCTAMCCGCTGAAAGAGGGGGAACCTGTTTATTTTT 771

Qy             61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
                  |||||||||||||||||| |||||||||||||| |||||| | |||||||||||||||||
Db            770 AGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 711
```

```
Qy            121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
                  |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db            710 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 651

Qy            181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
                  ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db            650 CCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 591

Qy            241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                    ||||||||||||||||||||||||| ||| |||||||||| |||||||||||||||||
Db            590 GCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAAT 531

Qy            301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
                   ||||||||||| |||            |||||||| ||| :||||||||||| |||||
Db            530 CGAAGCTGTAAAACTA------------CAAATGGAGCCCARGATGCAGTCCARGACTAG 483

Qy            361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
                   :||||||| || |||||||||||||||||| | |   |:|||||||||||||||||  |
Db            482 GDTCTACCGCGGCCCCCTGGACCGGCCTGCTSGCCCCCGMTCTGATGTTAATGACATCAA 423

Qy            421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
                  || ||||||||| | || ||||||| ||||||: || ||||||| | ||||||| | :||
Db            422 AGGCACCCCTCCTGGGGCAATCTCAGCTGCACMCCCTCTACTACGCCCCAATTCGGCRGG 363

Qy            481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
                  ||||||||:| || || ||| |||||||||||| ||:| ||||:||||||||||||||||
Db            362 AAGCAGTTRGGGCGGTCGTCGGCCAACCTCCCCGACRGCACTTRGGTTTTCCTGTTGAGA 303

Qy            541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
                   || ||||||||||||:||||| |||||||||||||||||||||||||||||| |:||||
Db            302 TGGGGGACTGAGAGACMGGACTGGCTGGATTTCCTAGGCTGACTAAGAATCCCTARGCCT 243

Qy            601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
                  : |||||||||||||| |||||| ||||||||: ||||||||| |||| |||||| || |
Db            242 RGCTGGGAAGGTGACCACATCCACCTTTAAACMCGGGGCTTGCCACTTSGCTCACCCCTG 183

Qy            661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGC- 719
                   ||||||:| | ||||||||||||||||:|: |||     | :: ::    ||  | ||
Db            182 CCCAATCRGGGGGCTCACTAAAATGCTAMTYCGGCCCCGSCSSSSSVCSSSGCCCTCGCG 123

Qy            720 CAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGC 779
                  |: :|:||||||||||||||||||||| ||| |||||| ||| ||||||||||| || |
Db            122 CMSYCMTCTATTGCCTGAGAGCACAGCMGGAGGGACAATGATCGGGATATAAACCCAMGT 63

Qy            780 ATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTT 839
                   |||  ||| ||  | ||  |||||||||||||||||||| ||      | |  || | |
Db             62 CTTCGCGCCGGCCCCGGCCCCCCCCTTTGGGTCCCCTCCCTTTCGTNNTGGGAGCTCTGT 3


RESULT 11
BX357208/c
LOCUS       BX357208                 758 bp    mRNA    linear   EST 08-APR-2004
DEFINITION  BX357208 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI022YJ18 3-PRIME, mRNA sequence.
ACCESSION   BX357208
VERSION     BX357208.2  GI:46305595
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 758)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30376125.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DI022DE09NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1. .758
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI022YJ18"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             45.8%;  Score 608.2;  DB 5;  Length 758;
  Best Local Similarity   88.1%;  Pred. No. 6.9e-173;
  Matches  665;  Conservative    9;  Mismatches   69;  Indels   12;  Gaps    1;

Qy          75 GTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAAATTTGAGATCGAA 134
               |||| |||||||||||||| |||||| | ||||||||||||||||||||| |||||||||
Db         758 GTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAA 699

Qy         135 TATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTCCTCAGCCAATGGA 194
               || || ||||||||||||| |||| |||||||| ||||||||||||||||||||||||||
Db         698 TACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGA 639

Qy         195 TGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTTTTACTCCTCTTTG 254
               ||||||||| ||||||||||||||||||||||||||||||||||||  ||||||||||||
Db         638 TGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTG 579

Qy         255 GACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGC 314
               ||||||||||||| ||| |||||||||| ||||||||||||||||| ||||||||||| |
Db         578 GACCCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAAAC 519

Qy         315 TACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGAC 374
               ||            |||||||| ||| |||||||||||| |||||| ||||||||  |||
Db         518 TA------------CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGAC 471

Qy         375 CCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCG 434
               ||||||||||||||||||| | | | |||||||||||||||||  ||| ||||||||| |
Db         470 CCCTGGACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTG 411

Qy         435 AGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCA 494
               ||||||||||| |||||||||| ||||||| | ||||||||| ||||||||||||||||
Db         410 AGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCG 351

Qy         495 GTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAG 554
               || ||| ||||||||||||||||| |||| |||||||||||||||| || ||||||||||
Db         350 GTCGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAG 291

Qy         555 ACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGA 614
               ||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||
Db         290 ACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGA 231

Qy         615 CCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGC 674
               || |||||| ||||||||| |||||||||||||||||||||||| |||||||||||||||
Db         230 CCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGC 171

Qy         675 TCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCC 734
               |||||||||||||||| ||||||| ||||||||||||| |:||||||||||||||||:||
Db         170 TCACTAAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAMTAGCCAATCATCTATTSCC 111

Qy         735 TGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAAC 794
               ::|||||||||| ||| |||||| ||| ||||||||||| |: ::   |  ||| || ||
Db         110 YSAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCMMSYCCCCGCGCCGGCCAC 51

Qy         795 AGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGG 829
```

```
                ||:||||||: ||||||||||||| |||||||||
Db          50 GGCMACCCCCYCTGGGTCCCCTCCCTTTGTATGGG 16


RESULT 12
BI087886
LOCUS       BI087886                 771 bp    mRNA    linear   EST 20-JUN-2001
DEFINITION  602852690F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4993894 5',
            mRNA sequence.
ACCESSION   BI087886
VERSION     BI087886.1  GI:14506216
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11015  row: d  column: 23
            High quality sequence stop: 762.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4993894"
                     /cell_line="MGC36"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_10"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5 kb. Library prepared by Life
                     Technologies."
ORIGIN

  Query Match             44.7%;  Score 593.4;  DB 4;  Length 771;
  Best Local Similarity   89.2%;  Pred. No. 2.2e-168;
  Matches  688;  Conservative    0;  Mismatches   68;  Indels   15;  Gaps    4;

Qy         107 GAGAAAGTTAAAGAAATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACT 166
               |||||||||||||||||| ||||||||||| || ||||||||||||| |||| |||||||
Db           1 GAGAAAGTTAAAGAAATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACT 60

Qy         167 GCACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTA 226
               | ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db          61 GGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTA 120

Qy         227 GCAGCTATAATATTTTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTT 286
               ||||||||||||||  ||||||||||||||||||||||||| ||| |||||||||| |||
Db         121 GCAGCTATAATATTGCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTTGTTAACTTT 180

Qy         287 GTCTCTTCCAGAATTGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATG 346
               |||||||||||||| ||||||||||| |||            |||||||| ||| |||||
Db         181 GTCTCTTCCAGAATCGAAGCTGTAAAACTA------------CAAATGGAGCCCAAGATG 228

Qy         347 CAGTCCATGACTAAAATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGAT 406
               ||||||| |||||| ||||||||  |||||||||||||||||||||| | | | ||||||
Db         229 CAGTCCAAGACTAAGATCTACCGCAGACCCCTGGACCGGCCTGCTAGCCCACGATCTGAT 288

Qy         407 GTTAATGACATTGAAGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACAC 466
               |||||||||||  ||| ||||||||| |||||||||||| |||||||||| ||||||| |
Db         289 GTTAATGACATCAAAGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGC 348
```

```
Qy    467 TCCAATTCAGTAGGAAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGG 526
           ||||||||| |||||||||||||||| || ||| | |||||||||||| || |||| ||
Db    349 CCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTC-GGCAACCTCCCCAAAAGCACTTAGG 407

Qy    527 TTTTCCTGT-TGAGAGGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTA 585
          ||||||||| ||||| || |||||||||||||||||||||||||||||||||||||||||
Db    408 TTTTCCTGTGTGAGATGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTA 467

Qy    586 AGAATCCCNAAGCCTANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAA 645
          |||||||| ||||||| |||||||||||||| |||||| ||||||||| |||||||||||
Db    468 AGAATCCCTAAGCCTAGCTGGGAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAA 527

Qy    646 CTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGA 705
          ||||||||||||| ||||||||||||||||||||||||||||||| ||||||| ||||||
Db    528 CTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGACAGGA 587

Qy    706 GGTA-AAGCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGG 764
          |||| ||| ||||||||||||||||||||||||||||||||| ||| |||||| ||| ||
Db    588 GGTACAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGG 647

Qy    765 GATATAAACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGT 824
          ||||| ||  || |  ||| |||| ||||| |||||||||||||||||||||||| ||||
Db    648 GATATTAAACCAAGTCTTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCCCTCCCTTTGT 707

Qy    825 ATGGGAGCTCTGTTTTCACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
          ||||||||||||| ||||  |||||||||||||||||||| ||||||||||
Db    708 ATGGGAGCTCTGTATTCATGCTATTTCACTCTATTAAATCTTGCAACTGCA 758

RESULT 13
BC035153
LOCUS       BC035153                4204 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens hypothetical protein FLJ22313, mRNA (cDNA clone
            IMAGE:5265109), with apparent retained intron.
ACCESSION   BC035153
VERSION     BC035153.1  GI:23272890
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4204)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 4204)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 73 Row: k Column: 22
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1..4204
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5265109"
                     /tissue_type="Brain, hippocampus"
                     /clone_lib="NIH_MGC_95"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
ORIGIN

  Query Match             44.5%;  Score 592;  DB 3;  Length 4204;
  Best Local Similarity   81.1%;  Pred. No. 9.9e-168;
  Matches  755;  Conservative    0;  Mismatches  128;  Indels   48;  Gaps    4;

Qy    400 CTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCT 459
          | ||||  |   ||| |   ||| ||||||| | ||||||||||||||||||||||||||
Db    887 CACTGAATTAGTTGAGAGACAAGGCACCCCTACGGAGGAAATCTCAACTGCACAACCCCT 946

Qy    460 ACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGT 519
          |||   | ||||||||| |||||||||||||||| || ||||   ||||||||||| || 
Db    947 ACTGTGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTGGTCA---AACCTCCCCAATAGC 1003

Qy    520 ACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGC 579
          |||||||||||||||||||||||| | |||||||||||||||||| ||||||||||||||
Db   1004 ACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAGAGACAGGACTAGTTGGATTTCCTAGGC 1063

Qy    580 TGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGC 639
          || ||||||||||| ||||||| |||||||  ||||| | |||| ||||||||| |||||
Db   1064 TGGCTAAGAATCCCTAAGCCTAGCTGGGAAATTGACCACGTCCACCTTTAAACACGGGGC 1123

Qy    640 TTGCAACTTAGCTCACACCCGACCAATC---------AGAGAGCTCACTAAAATGCTAAT 690
          |||||| |||||||||||||||||||||         |||||||||||||||||||||||
Db   1124 TTGCAATTTAGCTCACACCCGACCAATCAGGTAGTAAAGAGAGCTCACTAAAATGCTAAT 1183

Qy    691 CAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGA 750
           ||| ||||||||||||||||| | ||||||||||||||| ||||||||||||| | |||
Db   1184 TAGGGAAAAACAGGAGGTAAAGAAGTAGCCAATCATCTATCGCCTGAGAGCACAACAGGA 1243

Qy    751 AGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGG 810
           |||||| |||  |||||||||| |||||||||||||||||    || |||| |||||||
Db   1244 GGGACAATGATCAGGATATAAACCCAGGCATTCAAGCCAGCGGTGGCTACCCTCTTTGGG 1303

Qy    811 TCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCTATTAAATCATGCAA 870
          ||||||||| |||||||| ||||||||||||||||||||  |   |   |   ||
Db   1304 TCCCCTCCCTTTGTATGGAAGCTCTGTTTTCACTCTATTAAATCTTGCAATTGCA----- 1358

Qy    871 CTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACC 930
             |||| ||||||| ||||| | | |  |||||||||||||||| | || || ||||||
Db   1359 ---CACTTTTCTGGTACGTGTGTGTCACAGCTCAAGCTGAGCTTTCGCTCACCGTCCACC 1415

Qy    931 ACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGA 990
          |||||||| |||| | ||||||||||| | |||||||||||||||| ||
Db   1416 ACTGCTGTCTGCCGCTGTCACAGACCCACAGCTGACTTCCATCCCTCTG----------- 1464
```

```
Qy       991 GTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCATTGCCACTCCCGATCAGGCTAAAGG 1050
                              ||||||||||||  ||||||||  |||| |||| |||||||||
Db      1465 -----------------GATCCAGCGAGGCGCCCATTGCTGCTCCTGATCGGGCTAAAGG 1507

Qy      1051 CTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTCCTAATAGAACTGAACACTGG 1110
             |||||| ||||||||||| |||||||||||  |||| ||||||||  | || |||||| |
Db      1508 CTTGCCCTTGTTCCTGCAGGGCTAAGTGCCCAGGTTCGTCCTAATCTAGCTAAACACTAG 1567

Qy      1111 TCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCTTCTAATAGAGCTATAACACTCA 1170
             ||||||||||||| | |||||||||||||||||| ||||||||||||||||||||||||
Db      1568 TCACTGGGTTCCACGATTCTCTTCCATGACCCACAGCTTCTAATAGAGCTATAACACTCG 1627

Qy      1171 CCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGA 1230
             || | ||||||||||||||||||||||| |||  ||||||||||||||||||||||||||
Db      1628 CCACTTGGCCCAAGATTCCATTCCTTGGAATCCATGAGGCCAAGAACCCCAGGTCAGAGA 1687

Qy      1231 ANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGCCATTTTGGTAGCGGCCCACCAC 1290
             |    || ||||||||||| | |||||||||||    |||| |||| |||| ||  ||||
Db      1688 ACACAAGACTTGCCACCATCTCGGAAGTGGCCCGTCACCATCTTGGAAGCGACCTGCCAC 1747

Qy      1291 CATCTTGGGAGCTGTGGGAGCAAGGATCCCC 1321
             |||||||||| ||  ||||||||||| ||||
Db      1748 CATCTTGGGAACTTGGGGAGCAAGGACCCCC 1778


RESULT 14
CN272394
LOCUS       CN272394                 719 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000600005410 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN272394
VERSION     CN272394.1  GI:47288808
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 719   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1..719
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cell, retinoic acid and
                     mitogen-treated hES cell line H7"
                     /clone_lib="GRN_PRENEU"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from hES cell line H7 (p29) maintained in feeder-free
                     conditions. Embryoid bodies were generated in the presence
                     of all-trans retinoic and mitogens."
ORIGIN

  Query Match             42.2%;  Score 561;  DB 7;  Length 719;
  Best Local Similarity   85.4%;  Pred. No. 1.5e-158;
  Matches  657;  Conservative    0;  Mismatches   62;  Indels   50;  Gaps    1;

Qy        95 GGAATCATTACTGAGAAAGTTAAAGAAATTTGAGATCGAATATAATGTAGAGCAGAGGAC 154
             |||||| | ||||||||||||||||||||| ||||||||||| || |||||||||||||
Db         1 GGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAACGTAGAGCAGAGGAG 60
```

```
Qy       155 CTTCAAAACACTGCACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGACTCTCCCCTTC 214
             |||| |||||||| ||||||||||||||||||||||||||||||||||| ||||||||||
Db        61 CTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTC 120

Qy       215 TTAGGACCTCTAGCAGCTATAATATTTTTACTCCTCTTTGGACCCTGTATCTTCAACTTC 274
             ||||||||||||||||||||||||||  ||||||||||||||||||||||||| ||| ||
Db       121 TTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTGGACCCTGTATCTTTAACCTC 180

Qy       275 CTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTACAAATAGTTCTTCAAATG 334
             ||||||                                                  ||||
Db       181 CTTGTT--------------------------------------------------AATG 190

Qy       335 GAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCCCTGGACCGGCCTGCTAGA 394
             || ||| |||||||||||| |||||| ||||||||  ||||||||||||||||||||||
Db       191 GAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGGACCGGCCTGCTAGC 250

Qy       395 CTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAGGAAATCTCAACTGCACAA 454
             | | | |||||||||||||||||  ||| ||||||||| |||||||||||| ||||||||
Db       251 CCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAATCTCAGCTGCACAA 310

Qy       455 CCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCA 514
             || ||||||| | ||||||||| |||||||||||||||| || ||| |||||||||||||
Db       311 CCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCA 370

Qy       515 ACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGACAGGACTAGCTGGATTTCC 574
             |||| |||| |||||||||||||||| || ||||||||||||||||||||||||||||||
Db       371 ACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGACAGGACTAGCTGGATTTCC 430

Qy       575 TAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACCGCATCCATCTTTAAACAT 634
             ||||||||||||||||||| ||||||| |||||||||||||| |||||| |||||||||
Db       431 TAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACATCCACCTTTAAACAC 490

Qy       635 GGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAAATGCTAATCAGG 694
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||| |||
Db       491 GGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATGCTAATTAGG 550

Qy       695 CAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGA 754
             |||| ||||||||||||| ||||||||||||||||||||||||||||||||| ||| |||
Db       551 CAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCAGGAGGGA 610

Qy       755 CAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCC 814
             ||| ||| ||||||||||| || |  ||| |||| ||||| |||||||||||||||||||
Db       611 CAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGGCAACCCCCTTTGGGTCCC 670

Qy       815 CTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCTATTAAAT 863
             ||||| ||||||||||||||||||||||  |||||||||||||||||||
Db       671 CTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCTATTAAAT 719


RESULT 15
BX459153/c
LOCUS       BX459153                 900 bp    mRNA    linear   EST 05-MAY-2004
DEFINITION  BX459153 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE013YI20
            3-PRIME, mRNA sequence.
ACCESSION   BX459153
VERSION     BX459153.2  GI:47051796
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 900)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 22, 2003 this sequence version replaced gi:31021096.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DE013BE10NP1&c=4215.r.
FEATURES             Location/Qualifiers
     source          1. .900
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE013YI20"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             40.9%;  Score 544;  DB 5;  Length 900;
  Best Local Similarity   77.5%;  Pred. No. 2.4e-153;
  Matches  631;  Conservative   49;  Mismatches  120;  Indels   14;  Gaps    3;

Qy          1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAAGAGGGGGAACCTGTTTATTTTT 60
              |||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        802 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTT 743

Qy         61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
              |||||||||||||||||| |||||||||||||| |||||| | | |||||||||||||||||
Db        742 AGGGGAAGAATGCTGTTACTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 683

Qy        121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCT 180
              |||| ||||||||||| || ||||||||||||| |||| |||||||| ||||||||||||
Db        682 AATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCT 623

Qy        181 CCTCAGCCAATGGATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
              ||||||||||||||||||||||| ||||||| || ||:|||||||:||||||||||||||
Db        622 CCTCAGCCAATGGATGCCCTGGATTCTCCCCCTCCTASGACCTCTMGCAGCTATAATATT 563

Qy        241 TTTACTCCTCTTTGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAAT 300
                |||||||||||:||||||| ||||| ||| |||||| ||| |||||||||||||||||
Db        562 GCTACTCCTCTTTSGACCCTG-ATCTTTAACCTCCTTG-TAACTTTGTCTCTTCCAGAAT 505

Qy        301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAA 360
               ||||||||||| | |            |||||||| ||| |||||||| ||| :|| ||
Db        504 CGAAGCTGTAAAACCA------------CAAATGGAGCCCAAGATGCAGCCCAASACCAA 457

Qy        361 AATCTACCGTGGACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
               | ||||||  ||||||||||||:||| |||:| | | : |||||||| |||:|||   |
Db        456 SACCTACCGCAGACCCCTGGACCSGCCCGCTMGCCCACSATCTGATGTCAATSACACCAA 397

Qy        421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGG 480
              |: |||||| ||:||:|||||| || | |||||||| ||||||| | ||||||||| |:|
Db        396 ASGCACCCCCCCMGASGAAATCCCAGCCGCACAACCTCTACTACGCCCCAATTCAGCASG 337

Qy        481 AAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
              ||||||||||||| || :||  |||||| ||||:|||| |||| :|||| ||||||:|||
Db        336 AAGCAGTTAGAGCGGTCSTCSCCCAACCCCCCCMACAGCACTTASGTTTCCCTGTTSAGA 277

Qy        541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCT 600
               ||  : |||:|||:||||:|||||||::||  || :  |  || :: || ||| || |||
Db        276 TGGGSCACTSAGASACAGSACTAGCTSSATCMCCCMCCCCCACCMMCAACCCCCAACCCT 217

Qy        601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
              : |||||||||||||| || ||: |   |:||| ||:||  :| |||| ||||||:|| |
Db        216 MGCTGGGAAGGTGACCACACCCMCCCCCAMACACGGSGCCCSCCACTTCGCTCACVCCTG 157

Qy        661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
               ||||||||||||| :::  ||| |||::  ||||||| |||||||| |: | :||:|||
Db        156 SCCAATCAGAGAGCGSRSGGAAAGGCTRRCTAGGCAAAGACAGGAGGCARGGRRATMGCC 97

Qy        721 AATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCA 780
              :: |||: | |||||||||||||:|| ||| ||||:| |:  |||| :||||: || |
Db         96 MVCCATSCACTGCCTGAGAGCACRGCAGGAGGGACVATGMGCGGGAGRTAAASCCAAGTC 37

Qy        781 TTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCC 814
              : |  ||| || || ||  ||||||:|||  |||
Db         36 BCCGCGCCGGCTACGGCKCCCCCCTKTGGTCCCC 3


RESULT 16
BX953822/c
LOCUS       BX953822                1002 bp    mRNA    linear   EST 01-MAR-2004
DEFINITION  DKFZp781M064_r1 781 (synonym: hlcc4) Homo sapiens cDNA clone
            DKFZp781M064 5', mRNA sequence.
ACCESSION   BX953822
VERSION     BX953822.1  GI:43433374
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1002)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     EST (Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp781M064) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .1002
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp781M064"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="781 (synonym: hlcc4)"
                     /note="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             37.7%;  Score 500.6;  DB 5;  Length 1002;
  Best Local Similarity   87.6%;  Pred. No. 3.8e-140;
  Matches  641;  Conservative    0;  Mismatches   72;  Indels   19;  Gaps    8;

Qy        549 TGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGA 608
              |||||||||| ||||||||||||||||||||  |||||||||||| ||||||| ||||||
Db        922 TGAGAGACAGAACTAGCTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGA 863

Qy        609 AGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCA 668
              | ||||||||||||| ||||||||| |||||||||||||| ||||||||||    |
Db        862 A-GTGACCGCATCCACCTTTAAACACGGGGCTTGCAACTTGGCTCACACCCAGAGA---- 808

Qy        669 GAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCT 728
                   ||||||||||||||||| ||||||||||||||||||||| ||||||| |||||||
Db        807 -----CTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCGATCATCT 753

Qy        729 ATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCC 788
              |||||||||||||||||| |||||||||||||| ||||||||||  ||||||||| ||||
Db        752 ATTGCCTGAGAGCACAGC-GGAAGGACAAGGATCGGGATATAAAACCAGGCATTCGAGCC 694

Qy        789 AGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGAGCTCTGTTTTCACTCTAT 848
```

```
                |||||  |||||||||||||||||||||||| | ||||||||||||||||||||||||||
Db          693 AGCAATGGCAACCCCCTTTGGGTCCCCTCCCTTCGTATGGAGCTCTGTTTTCACTCTAT 634

Qy          849 TTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCT 908
                ||||| |||||||||| || ||||||||||||||||||| |||||| |||  |||| ||||
Db          633 TTCACACTATTAAATCTTGAAACTGCACTCTTCTGGTTCGTGTTTGTTACTGCTCGAGCT 574

Qy          909 GAGCTTTTGTTCGCCATCCACCACTGCTGTT-TGCCACCGTCACAGACCCGCTGCTGACT 967
                ||||||| | || |  |||||||||||||||| |||| ||||  ||||| ||||||||||||
Db          573 GAGCTTTCGCTCACTGTCCACCACTGCTGTTTTGCCGCCGTAGCAGACTCGCTGCTGACT 514

Qy          968 TCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTA-CCCAT 1026
                | ||| | |  ||||||||||| |||||||||||||||||||||||||||||| | ||| |
Db          513 TACATTCTTCCGGATCCAGCAGGGTGTCCACTGTGCTCCTGATCCAGCGAGGCACCCCTT 454

Qy         1027 TGCCACTCCCGATCAGGCTAAAGGCTTGCCA-TTGTTCCTGCATGGCTAAGTGCCTGGGT 1085
                |||| |||| |||| ||||||||||||||||| |||||||||||| ||||||||||||||||
Db          453 TGCCGCTCCAGATCGGGCTAAAGGCTTGCCATTTGTTCCTGCACGGCTAAGTGCCTGGGT 394

Qy         1086 TTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCACG 1145
                |||||||||| || |||||||||| |||||||||||||||||||| |||||| |||||||||
Db          393 TTGTCCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTGTCTTCCGTGACCCACG 334

Qy         1146 GCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATCTGT 1205
                 |||||||||||||| |||||||    |||||||||||||||||| |||||| | ||| ||
Db          333 ACTTCTAATAGAGCTGTAACACT----GCATGGCCCAAGATTCCGTTCCTTAGAATCCGT 278

Qy         1206 GAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCAT-TTGGGAAGTGGCCCA 1264
                ||||||||||||||||||||||||||   ||||||||||| ||| || ||||||||||||
Db          277 GAGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCATCATCTTTGGAAGTGGCCCA 218

Qy         1265 CTGCCATTTTGG 1276
                |  |||| || |
Db          217 CGACCATCTTTG 206

RESULT 17
AG113694/c
LOCUS       AG113694                 679 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-120G11.R, genomic survey sequence.
ACCESSION   AG113694
VERSION     AG113694.1  GI:16734213
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 679)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector     : pKS145
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1..679
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-120G11.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match            36.2%;  Score 480.8;  DB 9;  Length 679;
  Best Local Similarity  86.7%;  Pred. No. 3.5e-134;
  Matches  552;  Conservative    0;  Mismatches   83;  Indels    2;  Gaps    2;

Qy          673 GCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTG 732
                ||||  ||| |||||||| |||||||||||||||||||||| |||||||||||||||||||
Db          678 GCTCGGTAAGATGCTAATAAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTG 619

Qy          733 CCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCA 792
                | |||||||||||| |  | | | |||| | ||||||| | | | |||||| |||| |||
Db          618 CGTGAGAGCACAGTGTAAGGAAAAAGGGTCGGGATATTAGCCCGGGCATTCGAGCCGGCA 559

Qy          793 AC-AGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCT-GTTTTCACTCTATTT 850
                ||  ||| |||||||||||||||||||| ||||||| ||||||| |||||||||||||||
Db          558 ACGGGCAGCCCCCTTTGGGTCCCCTCCCTTTGTATGGCAGCTCTGGTTTTCACTCTATTT 499

Qy          851 CACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCTGA 910
                |||| ||||| ||| |||| |||||||||||||||||| |||||| |||  |||||| | ||
Db          498 CACTATATTAGATCTTGCAGCTGCACTCTTCTGGTCGGTGTTTGTTACAGCTCAAACAGA 439

Qy          911 GCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTTCC 970
                ||||||| | ||| ||| |||||||||||| || ||||| ||||||||   |||||||||
Db          438 GCTTTTGCTTGCCGTCCGCCACTGCTGTTTACCGCCGTCCCAGACCCGTCACTGACTTCC 379

Qy          971 ATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCATTGCC 1030
                |||||| ||||||| |||| |||| | |||||| ||||||||||||||| ||||||||| 
Db          378 ATCCCTCTGGATCCGGCAGGGTGTGCGCTGTGCTACTGATCCAGCGAGGCACCCATTGCT 319

Qy         1031 ACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTC 1090
                 ||||||||| ||||||||| |||||||||||||||||||||||||||||||||||| |||
Db          318 GCTCCCGATCGGGCTAAAGGGTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTCGTC 259

Qy         1091 CTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCTTC 1150
                 |||| || |||||||||| ||| ||| ||||||  |||||||||| ||||||||||||| 
Db          258 GTAATGGAGCTGAACACTAGTCGCTGAGTTCCACAGTTCTCTTCCGTGACCCACGGCTTG 199

Qy         1151 TAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAGGC 1210
                ||||||| || |||| ||||||||||||||||||||||||| |||||||| ||| |||||||
Db          198 TAATAGAACTGTAACGCTCACCGCATGGCCCAAGATTCCGTTCCTTGGAATCCGTGAGGC 139

Qy         1211 CAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGCCA 1270
                |||||||||||||||||||||   |||||||||||||||| | |||||||||||| ||||||
Db          138 CAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGAAGTGGCCCGCTGCCA 79

Qy         1271 TTTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGG 1307
                |||||| || ||||| ||||||||||||||||||| | |
Db           78 TTTTGGAAGTGGCCCGCCACCATCTTGGGAGCTCTCG 42

RESULT 18
BM918330/c
LOCUS       BM918330                 999 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6708649 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5747240
            5', mRNA sequence.
ACCESSION   BM918330
VERSION     BM918330.1  GI:19368709
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 999)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
```

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12773  row: b  column: 09
            High quality sequence stop: 685.
FEATURES             Location/Qualifiers
     source          1..999
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5747240"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_120"
                     /note="Organ: pooled pancreas and spleen; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of spleen and pancreas from 28 yo
                     male.  Library is oligo-dT primed and directionally cloned
                     (EcoRV  site is destroyed upon cloning).  Average insert
                     size 1.5 kb, insert size range 1-2.5 kb.  Library is
                     normalized and enriched for  full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 025. Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             36.2%;  Score 480.8;  DB 4;  Length 999;
  Best Local Similarity   83.1%;  Pred. No. 4e-134;
  Matches  621;  Conservative    0;  Mismatches  110;  Indels   16;  Gaps    6;

Qy        549 TGAGAGACAGGACTAGCT-GGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGG 607
              |||||||||||||||||| |||||||  |||| ||||||||||||| | |||||  | ||
Db        886 TGAGAGACAGGACTAGCTGGGATTTCTAAGGCGGACTAAGAATCCCTAGGCCTAGTTAGG 827

Qy        608 AAGGTGACCGCATCCATCTTTAAA-CATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 666
              |||||||||||||||| | ||||| |||||||||||||||||||||||||||||||||||
Db        826 AAGGTGACCGCATCCACCCTTAAACCATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 767

Qy        667 C---------AGAGAGCTCACTAAAATGCTAATCAGGC-AAAAACAGGAGGTAAAGCAAT 716
              |         ||||||||||||||||||| |||||| | ||||||||||||||||| | |
Db        766 CAGGTAGTAAAGAGAGCTCACTAAAATGCCAATCAGACAAAAAACAGGAGGTAAAGAACT 707

Qy        717 AGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCA 776
              | ||||||||||||||||||||||||||| |||| |||||| |||  |||||||||  ||
Db        706 AACCAATCATCTATTGCCTGAGAGCACAGTGGGAGGGACAATGATCAGGATATAAAGCCA 647

Qy        777 GGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTG 836
              ||||||| |||| ||||| || |||| |||||| ||||||||| ||||||||||||| ||
Db        646 GGCATTCCAGCCGGCAACGGCTACCCTCTTTGGCTCCCCTCCCTTTGTATGGGAGCTGTG 587

Qy        837 TTTTCACTCTATTTCACTCTATTAAATCATGCAACT---GCACTCTTCTGGTCCGTGTTT 893
              |||||||||||||  |   | | |   ||| |   |   |  |   | ||  ||||||||
Db        586 TTTTCACTCTATTAAATCTTGTAACTGCATACTCTTCTGGTCCGTGTTTGTACCGTGTTT 527

Qy        894 TTTATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAG 953
               |||  ||||||||| |||||| | |||   ||||||||||||||||||| | ||| |||
Db        526 GTTACCGCTCAAGCTAAGCTTTCGCTCGTTGTCCACCACTGCTGTTTGCCGCAGTCGCAG 467

Qy        954 ACCCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCA 1013
              |||||| ||||||||||| ||||  |||||  |||| |||||  ||||||||||||||||
Db        466 ACCCGCCGCTGACTTCCACCCCTCCGGATCTGGCAGGGTGTCTGCTGTGCTCCTGATCCA 407

Qy       1014 GCGAGGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCT 1073
              ||||||  |||| |||  ||||||||| |||| |||||||||| ||||||||||| ||| 
Db        406 GCGAGGCGCCCACTGCTGCTCCCGATCGGGCTGAAGGCTTGCCGTTGTTCCTGCACGGC- 348

Qy       1074 AAGTGCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTT 1133
              ||||||||||||| |||||||| ||  |||||||| ||| |||||||| | |||||||||
Db        347 AAGTGCCTGGGTTCGTCCTAATTGAGATGAACACTAGTCGCTGGGTTCTACGGTTCTCTT 288
```

```
Qy       1134 CCATGACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTC 1193
              |||||||||||| |||| ||||||||||||||| ||||||||||||||||||||||||||
Db        287 CCATGACCCACGTCTTCCAATAGAGCTATAACATTCACCGCATGGCCCAAGATTCCATTC 228

Qy       1194 CTTGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGG 1253
              ||||| ||| ||| |||||||||||||||||||||||| | |||||||||  |||| | |
Db        227 CTTGGAATCCGTGGGGCCAAGAACCCCAGGTCAGAGAACGAGAGGCTTGCTGCCATCTTG 168

Qy       1254 GAAGTGGCCCACTGCCATTTTGGTAGC 1280
              ||||   ||| |  |||| |||| |||
Db        167 GAAGCAACCCGCCACCATGTTGGGAGC 141


RESULT 19
AG134524
LOCUS       AG134524                 683 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-147I10.R, genomic survey sequence.
ACCESSION   AG134524
VERSION     AG134524.1  GI:16664202
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 683)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: M13Rev
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1..683
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-147I10.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             35.8%;  Score 475.2;  DB 9;  Length 683;
  Best Local Similarity   89.6%;  Pred. No. 1.8e-132;
  Matches  554;  Conservative    0;  Mismatches   60;  Indels    4;  Gaps    4;

Qy        666 TCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCA 725
              ||  ||||||||||||||||||||| ||||||||||||||||||||| ||||||||||||
Db         58 TCGNAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCA 117

Qy        726 TCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAA 785
              |||||||||||||||||||| ||||||||||||||| ||||||||||| |||||||||||
Db        118 TCTATTGCCTGAGAGCACAGTGGGAAGGACAAGGATCGGGATATAAACCCAGGCATTCAA 177

Qy        786 GCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTC 845
              ||||||||| |||||||||||||||| ||||||| |||||||||||||||||||||||||
Db        178 GCCAGCAACGGCAACCCCCTTTGGGT-CCCTCCCCTTGTATGGGAGCTCTGTTTTCACTC 236

Qy        846 TATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAA 905
              ||||||||||||||||||| |||||||||||||||||||||||||||| ||| ||||| |
Db        237 TATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCGTGTTTGTTACGGCTCGA 296

Qy        906 GCTGAGCTTTTGTTCGCCATCCACCACTGCTG-TTTGCCACCGTCACAGACCCGCTGCTG 964
              ||||||||||   ||||||||||||||||||| |||||| ||||| |||||| || ||||
Db        297 GCTGAGCTTTCTCTCGCCATCCACCACTGCTGTTTTGCCGCCGTCGCAGACCGGCCGCTG 356

Qy        965 ACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCC 1024
              ||||||||||||   |||||||||| |||||| ||||||||||||||||||||||  |||
Db        357 ACTTCCATCCCTCCAGATCCAGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGCGCCC 416

Qy       1025 ATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGG 1084
              |||||| |||| |||| ||||||||||||||||||||||||||| ||||||||||||| |
Db        417 ATTGCCGCTCCTGATCGGGCTAAAGGCTTGCCATTGTTCCTGCACGGCTAAGTGCCTGTG 476

Qy       1085 TTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCAC 1144
              || ||||||||  | ||||||||| |||||||||||||| ||||||||||||||| ||||
Db        477 TTCGTCCTAATCAAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCCATGAACCAC 536

Qy       1145 GGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCT-TGGTATCT 1203
              | ||||||||||||||||||||||||| |||||||| ||||||||||| ||  || |||
Db        537 GACTTCTAATAGAGCTATAACACTCACTGCATGGCCTAAGATTCCATTTCTGGGGAATCC 596

Qy       1204 GTGAGGCCAA-GAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCC 1262
               ||||||||| ||||||||||||||| ||   ||||||||||||||  | ||||||||||
Db        597 ATGAGGCCAATGAACCCCAGGTCAGAAAACACGAGGCTTGCCACCACCTTGGAAGTGGCC 656

Qy       1263 CACTGCCATTTTGGTAGC 1280
              |||  |||| | || | |
Db        657 CACCACCATCTGGGGAAC 674


RESULT 20
AG049953
LOCUS       AG049953                 669 bp    DNA     linear   GSS 02-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-030D19.F, genomic survey sequence.
ACCESSION   AG049953
VERSION     AG049953.1  GI:16586845
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 669)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector     : pKS145
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1..669
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-030D19.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             35.7%;  Score 474;  DB 9;  Length 669;
  Best Local Similarity   86.8%;  Pred. No. 4.1e-132;
  Matches  548;  Conservative    0;  Mismatches   71;  Indels   12;  Gaps    2;

Qy        669 GAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCT 728
              ||||||||||||||||||| || |||||||| |||||||||||| |||||||||||||||
Db         49 GAGAGCTCACTAAAATGCTGATTAGGCAAAAGCAGGAGGTAAAGAAATAGCCAATCATCT 108

Qy        729 ATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCC 788
              |||||||||||||||||| ||| |||||| ||| ||||||||||| ||||||||  |||
Db        109 ATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCATTGCAGCT 168

Qy        789 AGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTAT 848
               ||||| || |||  |||||||||||||||| ||||||| |||||||||          |
Db        169 GGCAACGGCTACCTTCTTTGGGTCCCCTCCCTTTGTATGAGAGCTCTGT----------T 218

Qy        849 TTCACTCTATTAAATCATGCAACTGCACTC--TTCTGGTCCGTGTTTTTTATGGCTCAAG 906
              |||||||||||||||| |||||||||||||  || || ||| ||||| |||  |||| ||
Db        219 TTCACTCTATTAAATCTTGCAACTGCACTCTTTTATGTTCCATGTTTGTTACAGCTCGAG 278

Qy        907 CTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGAC 966
               |||||||||| | ||| ||||||| ||||||||||  ||||| ||||||| | ||||||
Db        279 GTGAGCTTTTGATTGCCGTCCACCAGTGCTGTTTGCTGCCGTCGCAGACCCACCGCTGAC 338

Qy        967 TTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCAT 1026
              |||||||||| ||||||  |||| |||||| ||||||||||||||||||||||| |||||
Db        339 TTCCATCCCTCTGGATCTGGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGTGCCCAT 398

Qy       1027 TGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTT 1086
              |||  |||| |||| || |||||||||||||||||||||||| |||||||||||||||||
Db        399 TGCTGCTCCTGATCGGGATAAAGGCTTGCCATTGTTCCTGCAAGGCTAAGTGCCTGGGTT 458

Qy       1087 TGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCACGG 1146
               |||||||| || ||||||||| |||||||||||||||||||||||||| |||| || ||
Db        459 CGTCCTAATCGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCCGTGACTCATGG 518

Qy       1147 CTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTG 1206
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||| | |
Db        519 CTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGAATCCGGG 578

Qy       1207 AGGCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACT 1266
              |||||||||||||||||||||||||   ||||||||||||||| | |||||||| ||||
Db        579 AGGCCAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATCTTGGAAGTGGACCACC 638

Qy       1267 GCCATTTTGGTAGCGGCCCACCACCATCTTG 1297
               ||||||||| ||| |||| ||||||| |||
Db        639 ACCATTTTGGAAGCAGCCCGCCACCATGTTG 669


RESULT 21
AG116971
LOCUS       AG116971                 681 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-124K18.F, genomic survey sequence.
ACCESSION   AG116971
VERSION     AG116971.1  GI:16737490
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 681)
```

```
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE     Direct Submission
JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
          Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
          was generated during the R&D process and may have higher chance of
          clone tracking errors.
           PRIMERS
              Sequencing: -21M13
           LIBRARY
              Vector    : pKS145
              R.Site 1  : SacI
              R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1..681
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-124K18.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             35.6%;  Score 473;  DB 9;  Length 681;
  Best Local Similarity   88.8%;  Pred. No. 8.2e-132;
  Matches  538;  Conservative    0;  Mismatches   56;  Indels   12;  Gaps    2;

Qy        671 GAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTAT 730
              ||||||||||||| |||||| ||||||||||||||||||||| |||||||||||||||||
Db         86 GAGCTCACTAAAAAGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTAT 145

Qy        731 TGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAG 790
              |||||||||||||||| ||| |||||| ||||||||||||||| ||||||||| ||||||
Db        146 TGCCTGAGAGCACAGCAGGAGGGACAATGATTGGGATATAAACCCAGGCATTCGAGCCAG 205

Qy        791 CAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTT 850
              |||| || |||| |||||||||||||||| |||||||||||||||||          |||
Db        206 CAACGGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGT----------TTT 255

Qy        851 CACTCTATTAAATCATGCAACTGCA--CTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCT 908
              |||||||||||||| ||||||||||  |||||||||||||||||| ||||||||||||||
Db        256 CACTCTATTAAATCTTGCAACTGCACTCTCTTCTGGTCCGTGTTTGTTATGGCTCAAGCT 315

Qy        909 GAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTT 968
              ||||||| | | ||||||||||||||||||||||| ||| | |||||||||  |||||||
Db        316 GAGCTTTCGCTTGCCATCCACCACTGCTGTTTGCCGCCGCCGCAGACCCGCCTCTGACTT 375

Qy        969 CCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCATTG 1028
              ||||||||   ||||  |||| |||||||||||||||||||||||||||||  |||||||
Db        376 CCATCCCTCCAGATCTGGCAGGGTGTCCACTGTGCTCCTGATCCAGCGAGGCGCCCATTG 435

Qy       1029 CCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTG 1088
              || ||||||||  |||||||||||||||||||||||||||||||||| ||||  ||||
Db        436 CCGCTCCCGATAGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAACTGCCCAGGTTCA 495

Qy       1089 TCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCT 1148
              ||||||| || ||||||||| || || || ||||||||||||||||| ||||||||||||
Db        496 TCCTAATCGAGCTGAACACTAGTTACCGGCTTCCATGGTTCTCTTCCGTGACCCACGGCT 555

Qy       1149 TCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAG 1208
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||||
Db        556 TCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGAATCCGTGAG 615

Qy       1209 GCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGC 1268
              |||||||||||||||||||||||    |||||||||||||| | |||||||||| ||  |
Db        616 GCCAAGAACCCCAGGTCAGAGAACACAAGGCTTGCCACCATCTTGGAAGTGGCCAACCAC 675

Qy       1269 CATTTT 1274
```

```
              ||||||
Db        676 CATTTT 681


RESULT 22
AG096341
LOCUS       AG096341                 710 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-097K23.F, genomic survey sequence.
ACCESSION   AG096341
VERSION     AG096341.1  GI:16716858
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 710)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1..710
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-097K23.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             35.3%;  Score 469.6;  DB 9;  Length 710;
  Best Local Similarity   85.1%;  Pred. No. 9e-131;
  Matches  561;  Conservative    0;  Mismatches   65;  Indels   33;  Gaps    2;

Qy        671 GAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTAT 730
              |||||||||||||||||||| ||||||||||||||||||||| |||||||||||||| ||
Db         75 GAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCCAT 134

Qy        731 TGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAG 790
              ||||||||||||||| |||| |||||| ||   |||||||||| ||||||||| |||| |
Db        135 TGCCTGAGAGCACAGTGGGATGGACAATGACCAGGATATAAACCCAGGCATTCCAGCCTG 194

Qy        791 CAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTT 850
              |||| |||||||||||||| ||||||| | |||||||||||||||||          |||
Db        195 CAACGGCAACCCCCTTTGGTTCCCCTCTCTTTGTATGGGAGCTCTGT----------TTT 244

Qy        851 CACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCTGA 910
              ||||||||  |||| |||||||||||||||||||| ||||||| ||| ||||||||||||
Db        245 CACTCTATGCAATCTTGCAACTGCACTCTTCTGGTTCGTGTTTGTTACGGCTCAAGCTGA 304

Qy        911 GCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTTCC 970
              ||||||| || |||||||||||||||||||||| ||||  |||||||   ||||||||||
Db        305 GCTTTTGCTCACCATCCACCACTGCTGTTTGCCGCCGTTGCAGACCCATCGCTGACTTCC 364
```

```
Qy        971 ATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCATTGCC 1030
              ||||||   ||||  |||| ||||||||||||||||||||||||||||| ||||||||||
Db        365 ATCCCTCCAGATCTGGCAGGGTGTCCACTGTGCTCCTGATCCAGCGAGGCACCCATTGCC 424

Qy       1031 ACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTC 1090
               ||||||||| ||||||||||||||||||||||||||| ||||||||||||||||| |||
Db        425 GCTCCCGATCGGGCTAAAGGCTTGCCATTGTTCCTGCACGGCTAAGTGCCTGGGTTCGTC 484

Qy       1091 CTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCTTC 1150
              |||||  | ||||||||| |||||||||||||||||||||||||||||||||| ||||||
Db        485 CTAATCAAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCCATGACCCATGGCTTC 544

Qy       1151 TAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAGGC 1210
              |||||||||||||||||||||||||||||||||||||||||||||||| ||| |||||||
Db        545 TAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGAATCCGTGAGGC 604

Qy       1211 CAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGCCA 1270
              |||||||||||||||||||||   ||| ||                        |  |||
Db        605 CAAGAACCCCAGGTCAGAGAACACGAGACT-----------------------GCCACCA 641

Qy       1271 TTTGGTAGCGGCCCACCACCATCTTGGGAGCTGTGGGAGCAAGGATCCCCCAGTAACA 1329
              | |||| ||| |||  |||||||||||| |||| |  ||||||||| ||||||||||||
Db        642 TCTTGGAAGCAGCCTGCCACCATCTTGGAAGCTCTATGAGCAAGGACCCCCCAGTAACA 700


RESULT 23
BC030968
LOCUS       BC030968                1160 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens cDNA clone IMAGE:4724433, with apparent retained
            intron.
ACCESSION   BC030968
VERSION     BC030968.1  GI:22658419
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1160)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1160)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 41 Row: m Column: 10
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1..1160
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4724433"
                     /tissue_type="Placenta"
                     /clone_lib="NIH_MGC_79"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
ORIGIN

  Query Match             35.2%;  Score 468.2;  DB 3;  Length 1160;
  Best Local Similarity   90.1%;  Pred. No. 2.8e-130;
  Matches  500;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy        321 TAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCCCTGG 380
              ||   || |||||||| ||| |||||||||||| |||||| ||||||||  |||||||||
Db        577 TAAAACTACAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGG 636

Qy        381 ACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAGGAAA 440
              ||||||||||||| | | | |||||||||||||||||  ||| ||||||||| |||||||
Db        637 ACCGGCCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAA 696

Qy        441 TCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGTTGTC 500
              ||||| |||||||||| ||||||| | ||||||||| |||||||||||||||| || |||
Db        697 TCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTC 756

Qy        501 AGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGACAGGA 560
               ||||||||||||||||| |||| |||||||||||||||| || ||||||||||||||||
Db        757 GGCCAACCTCCCCAACAGCACTTAGGTTTTCCTGTTGAGATGGGGGACTGAGAGACAGGA 816

Qy        561 CTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGACCGCAT 620
              ||||||||||||||||||||||||||||||||| ||||||| |||||||||||||| |||
Db        817 CTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACAT 876

Qy        621 CCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTCACTA 680
              ||| ||||||||| |||||||||||||||||||||||| |||||||||||||||||||||
Db        877 CCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTA 936

Qy        681 AAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTGAGAG 740
              |||||||||| ||||||| ||||||||||||| |||||||||||||||||||||||||||
Db        937 AAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAG 996

Qy        741 CACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAGCAAC 800
              |||||| ||| |||||| ||| ||||||||||| || |  ||| |||| ||||| |||||
Db        997 CACAGCAGGAGGGACAATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGGCAAC 1056

Qy        801 CCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCTATTA 860
              ||||||||||||||||||| ||||||||||||||||||||||  ||||||||||||||||
Db       1057 CCCCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCTATTA 1116

Qy        861 AATCATGCAACTGCA 875
              |||| ||||||| ||
Db       1117 AATCTTGCAACTACA 1131


RESULT 24
BQ437925
LOCUS       BQ437925                 893 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AGENCOURT_7917208 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6161436
```

```
            5', mRNA sequence.
ACCESSION   BQ437925
VERSION     BQ437925.1  GI:21177001
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13513  row: 1  column: 13
            High quality sequence stop: 650.
FEATURES             Location/Qualifiers
     source          1..893
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6161436"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
ORIGIN

  Query Match             35.0%;  Score 465.2;  DB 5;  Length 893;
  Best Local Similarity   80.4%;  Pred. No. 2.1e-129;
  Matches  599;  Conservative    0;  Mismatches  101;  Indels   45;  Gaps    3;

Qy        586 AGAATCCCNAAGCCTANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAA 645
              |||||||| ||||||| |||||||  ||||| | |||| ||||||||| |||||||||||
Db          1 AGAATCCCTAAGCCTAGCTGGGAAATTGACCACGTCCACCTTTAAACACGGGGCTTGCAA 60

Qy        646 CTTAGCTCACACCCGACCAATC---------AGAGAGCTCACTAAAATGCTAATCAGGCA 696
               |||||||||||||||||||||         ||||||||||||||||||||||| ||| |
Db         61 TTTAGCTCACACCCGACCAATCAGGTAGTAAAGAGAGCTCACTAAAATGCTAATTAGGGA 120

Qy        697 AAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACA 756
              |||||||||||||||| | ||||||||||||||| ||||||||||||| | ||| |||||
Db        121 AAAACAGGAGGTAAAGAAGTAGCCAATCATCTATCGCCTGAGAGCACAACAGGAGGGACA 180

Qy        757 AGGATTGGGATATAAACTCAGGCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCT 816
              | |||  |||||||||| |||||||||||||||||    || |||| |||||||||||||
Db        181 ATGATCAGGATATAAACCCAGGCATTCAAGCCAGCGGTGGCTACCCTCTTTGGGTCCCCT 240

Qy        817 CCCATTGTATGGGAGCTCTGTTTTCACTCTATTTCACTCTATTAAATCATGCAACTGCAC 876
              ||| |||||||| ||||||||||||||||||||  |   |   |   ||        |||
Db        241 CCCTTTGTATGGAAGCTCTGTTTTCACTCTATTAAATCTTGCAATTGCA--------CAC 292

Qy        877 TCTTCTGGTCCGTGTTTTTTATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCT 936
              | ||||||| ||||| | | |  |||||||||||||||| | || || ||||||||||||
Db        293 TTTTCTGGTACGTGTGTGTCACAGCTCAAGCTGAGCTTTCGCTCACCGTCCACCACTGCT 352

Qy        937 GTTTGCCACCGTCACAGACCCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCC 996
              || |||| | ||||||||||| | |||||||||||||||| |||                
Db        353 GTCTGCCGCTGTCACAGACCCACAGCTGACTTCCATCCCTCTGG---------------- 396

Qy        997 ACTGTGCTCCTGATCCAGCGAGGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCC 1056
                          |||||||||||  ||||||||  |||| |||| |||||||||||||||
Db        397 ------------ATCCAGCGAGGCGCCCATTGCTGCTCCTGATCGGGCTAAAGGCTTGCC 444
```

```
Qy       1057 ATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTG 1116
               ||||||||||| |||||||||||  |||| ||||||||  | || |||||| |||||||
Db        445 CTTGTTCCTGCAGGGCTAAGTGCCCAGGTTCGTCCTAATCTAGCTAAACACTAGTCACTG 504

Qy       1117 GGTTCCATGGTTCTCTTCCATGACCCACGGCTTCTAATAGAGCTATAACACTCACCGCAT 1176
              ||||||| | |||||||||||||||||| |||||||||||||||||||||||| |||| |
Db        505 GGTTCCACGATTCTCTTCCATGACCCACAGCTTCTAATAGAGCTATAACACTCGCCGCTT 564

Qy       1177 GGCCCAAGATTCCATTCCTTGGTATCTGTGAGGCCAAGAACCCCAGGTCAGAGAANGTGA 1236
              |||||||||||||||||||||| |||  |||||||||||||||||||||||||||    |
Db        565 GGCCCAAGATTCCATTCCTTGGAATCCATGAGGCCAAGAACCCCAGGTCAGAGAACACAA 624

Qy       1237 GGCTTGCCACCATTTGGGAAGTGGCCCACTGCCATTTTGGTAGCGGCCCACCACCATCTT 1296
              | ||||||||||| | |||||||||||    |||| |||| |||| ||  ||||||||||
Db        625 GACTTGCCACCATCTCGGAAGTGGCCCGTCACCATCTTGGAAGCGACCTGCCACCATCTT 684

Qy       1297 GGGAGCTGTGGGAGCAAGGATCCCC 1321
              |||| ||  ||||||||||| ||||
Db        685 GGGAACTTGGGGAGCAAGGACCCCC 709

RESULT 25
AG072852
LOCUS       AG072852                 670 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-064H14.R, genomic survey sequence.
ACCESSION   AG072852
VERSION     AG072852.1  GI:16624654
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 670)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector     : pKS145
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1..670
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-064H14.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             34.3%;  Score 456.2;  DB 9;  Length 670;
  Best Local Similarity   86.4%;  Pred. No. 1e-126;
  Matches  529;  Conservative    0;  Mismatches   74;  Indels    9;  Gaps    2;

Qy        669 GAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCT 728
              ||||||||||||||| |||||| ||||||||||||||||||||| ||||||||||||| |
Db          57 GAGAGCTCACTAAAAAGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATTT 116

Qy         729 ATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCC 788
               |||||||||||||||||| ||| |||||| ||| ||||||||||| ||||||||  ||||
Db         117 ATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAAACCCAGGCATTGGAGCC 176

Qy         789 AGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTAT 848
               |||||| || |||| |||||||| ||||||| ||||||||||||||| ||||||||||||
Db         177 AGCAACGGCTACCCTCTTTGGGTTCCCTCCCTTTGTATGGGAGCTCTTTTTTCACTCTAT 236

Qy         849 TTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCT 908
               |  |   |   |   |        || ||||||||||| | |||| |||  |||||||||
Db         237 TAAATCTTGCAACTGC--------GCTCTCTTCTGGTCTGCGTTTGTTACAGCTCAAGCT 288

Qy         909 GAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTT 968
               ||| ||||| | ||  |||||||||||||||| || ||||| ||||||||||||||| ||
Db         289 GAG-TTTTGCTGGCTGTCCACCACTGCTGTTTTCCGCCGTCGCAGACCCGCTGCTGAGTT 347

Qy         969 CCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCATTG 1028
               ||||||||  ||||||| ||| ||||||||||||||||||||||||| |||  |||||||
Db         348 CCATCCCTCCGGATCCACCAGGGTGTCCACTGTGCTCCTGATCCAGCAAGGCGCCCATTG 407

Qy        1029 CCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTG 1088
               ||||||| |||| ||||||||||||||||||||||||||| || |||||||||||||| |
Db         408 CCACTCCTGATCGGGCTAAAGGCTTGCCATTGTTCCTGCACGGTTAAGTGCCTGGGTTCG 467

Qy        1089 TCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCT 1148
               ||||||| || ||||||||| |||||||||||||| |||||||||||||||||||| |||
Db         468 TCCTAATCGAGCTGAACACTAGTCACTGGGTTCCACGGTTCTCTTCCATGACCCACAGCT 527

Qy        1149 TCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAG 1208
               ||||||||||||||||||||||| |||||||||||||||||||||||||| ||| |||||
Db         528 TCTAATAGAGCTATAACACTCACTGCATGGCCCAAGATTCCATTCCTTGGAATCCGTGAG 587

Qy        1209 GCCAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGC 1268
               |||||||||||||||||||||||   |||||||||||||||||  |||| ||| | ||||
Db         588 GCCAAGAACCCCAGGTCAGAGAATACGAGGCTTGCCACCATTTTAGAAGCGGCTCGCTGC 647

Qy        1269 CATTTTGGTAGC 1280
               ||| |||  | |
Db         648 CATCTTGCGATC 659


RESULT 26
AG099717
LOCUS       AG099717                712 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-102B07.F, genomic survey sequence.
ACCESSION   AG099717
VERSION     AG099717.1  GI:16720234
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 712)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
            LIBRARY
                Vector     : pKS145
                R.Site 1   : SacI
                R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1..712
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-102B07.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             34.2%;  Score 455;  DB 9;  Length 712;
  Best Local Similarity   86.2%;  Pred. No. 2.5e-126;
  Matches  529;  Conservative    0;  Mismatches   76;  Indels    9;  Gaps    2;

Qy         671 GAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATCTAT 730
               |||||||||||||||||||| | ||||||||||||||||||| |||||||||||||||||
Db         107 GAGCTCACTAAAATGCTAATTA-GCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTAT 165

Qy         731 TGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAGCCAG 790
               |||||||||||| ||||||| |||||| ||| ||||||||||| ||||||||| ||||||
Db         166 TGCCTGAGAGCATAGCGGGAGGGACAATGATCGGGATATAAACCCAGGCATTCCAGCCAG 225

Qy         791 CAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCTATTT 850
               |||  || |||| |||||||||||||||| |||||||||||||||||||||||||||||
Db         226 CAAGGGCTACCCTCTTTGGGTCCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTATT- 284

Qy         851 CACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAGCTGA 910
                       ||  ||         | ||||||||| | || ||| ||||||||| ||||||
Db         285 -------ATATCTCGCAACTGCACTCTCTTCTGGCCTGTATTTGTTATGGCTCTAGCTGA 337

Qy         911 GCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGACCCGCTGCTGACTTCC 970
               ||||||| | ||| ||||||||||||||||||  || |||||||||||||||||||||||
Db         338 GCTTTTGCTAGCCGTCCACCACTGCTGTTTGCTGCCATCACAGACCCGCTGCTGACTTCC 397

Qy         971 ATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCATTGCC 1030
               |||||| ||||||| |||| |||||| ||||||||||||||||| ||| |||||||||||
Db         398 ATCCCTCTGGATCCGGCAGGGTGTCCTCTGTGCTCCTGATCCAGTGAGATACCCATTGCC 457

Qy        1031 ACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGTTTGTC 1090
               |||||||||  |||||||| |||||||||||||| |||||||||||||||  || |  ||
Db         458 ACTCCCGATTGGGCTAAAGACTTGCCATTGTTCCCGCATGGCTAAGTGCCCAGGCTCATC 517

Qy        1091 CTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCTTC 1150
               ||||| || ||||||||| ||||||||||||||  |||||||||||||| ||| ||||||
Db         518 CTAATCGAGCTGAACACTAGTCACTGGGTTCCACAGTTCTCTTCCATGATCCATGGCTTC 577

Qy        1151 TAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTATCTGTGAGGC 1210
               |||||||||||||||||| || |||||||| |||||  |||||||||| | |||||||||
Db         578 TAATAGAGCTATAACACTTACTGCATGGCCGAAGATCTCATTCCTTGGAACCTGTGAGGC 637

Qy        1211 CAAGAACCCCAGGTCAGAGAANGTGAGGCTTGCCACCATTTGGGAAGTGGCCCACTGCCA 1270
               |||||||||||||||||||||   ||||||||||||||||| |||| ||||||||  |||
Db         638 CAAGAACCCCAGGTCAGAGAACACGAGGCTTGCCACCATTTTGGAAATGGCCCACCACCA 697

Qy        1271 TTTTGGTAGCGGCC 1284
               | |||| || ||||
Db         698 TCTTGGAAGAGGCC 711


RESULT 27
AG121669/c
LOCUS       AG121669                689 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-130M15.F, genomic survey sequence.
ACCESSION   AG121669
VERSION     AG121669.1  GI:16650834
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 689)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector     : pKS145
                R.Site 1   : SacI
                R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-130M15.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             33.0%;  Score 439.2;  DB 9;  Length 689;
  Best Local Similarity   87.0%;  Pred. No. 1.6e-121;
  Matches  507;  Conservative    0;  Mismatches   66;  Indels   10;  Gaps    2;

Qy        257 CCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAAGCTA 316
              || || ||||| ||| ||||||||||| |||||||||||||||| ||||||||||| | |
Db        666 CCNTGGATCTTTAACGTCCTTGTTAAGCTTGTCTCTTCCAGAATCGAAGCTGTAAAACCA 607

Qy        317 CAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGGACCC 376
               ||    ||||||||||||| | ||  ||||||||||||||||| ||||||| |||||||
Db        606 TAATAGTTTCTTCAAATGGAGCACCTTATGCAGTCCATGACTAAGATCTACCATGGACCC 547

Qy        377 CTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCCCGAG 436
              ||||||| ||||||||| |  |||||||||||||||||||| |||| ||||||||| |||
Db        546 CTGGACCAGCCTGCTAGCCCATGCTCTGATGTTAATGACATCGAAGGCACCCCTCCTGAG 487

Qy        437 GAAATCTCAACTGCAC-AACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAG 495
              |||||||||||||||| ||||| |||||||||||| ||||| |||||||||||||||| |
Db        486 GAAATCTCAACTGCACAAACCCTTACTACACTCCAGTTCAGCAGGAAGCAGTTAGAGCGG 427

Qy        496 TTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAGAGA 555
              | ||| ||||||||||||||  | |||||||||||||||||||||||| |||||||||||
Db        426 TCGTCGGCCAACCTCCCCAATGGCACTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGA 367

Qy        556 CAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGTGAC 615
              |||||||| |||||||||||||||  |||||||||||| ||||||| |||||||||||||
Db        366 CAGGACTATCTGGATTTCCTAGGCCAACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 307

Qy        616 CGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC-------- 667
              | ||| ||  ||||||||  ||||||||||||||||||||||||||| ||||
Db        306 CACATTCACTTTTAAACACAGGGCTTGCAACTTAGCTCACACCCGACGAATCAGGTAGTA 247

Qy        668 -AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCAT 726
               |||| |||||||||||||||||| ||||||||||||||||||||| |||||||||| ||
Db        246 AAGAGGGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATAAT 187

Qy        727 CTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAG 786
```

```
               ||||||||||||| |||||||||||||||| ||| |||||||||||||| |||||| ||
Db        186 TTATTGCCTGAGAGTACAGCGGGAAGGACAATGATCGGGATATAAACTCATGCATTCGAG 127

Qy        787 CCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGG 829
              ||| |||  || |||| |||||||||||||||| ||||   ||
Db        126 CCACCAATGGCTACCCTCTTTGGGTCCCCTCCCTTTGTTATGG 84


RESULT 28
BH149565/c
LOCUS       BH149565                 921 bp    DNA     linear   GSS 27-AUG-2001
DEFINITION  ENTQG48TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION   BH149565
VERSION     BH149565.1  GI:15310303
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 40
            High quality sequence stop: 567.
FEATURES             Location/Qualifiers
     source          1. .921
                     /organism="Entamoeba histolytica"
                     /mol_type="genomic DNA"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
ORIGIN

  Query Match             33.0%;  Score 438.8;  DB 8;  Length 921;
  Best Local Similarity   83.8%;  Pred. No. 2.3e-121;
  Matches  550;  Conservative    0;  Mismatches   84;  Indels   22;  Gaps    4;

Qy        547 ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGG 606
              | |||||||||||||||||||||| || |||||  |||||||||||  |||| || ||||
Db        651 AGTGAGAGACAGGACTAGCTGGATATCTTAGGCCTACTAAGAATCCTTAAGCTTAGCTGG 592

Qy        607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 666
              ||||||||| ||||||| || |||||| ||||||||||||||||||||||||||||||||
Db        591 GAAGGTGACTGCATCCACCTATAAACACGGGGCTTGCAACTTAGCTCACACCCGACCAAT 532

Qy        667 C---------AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATA 717
              |         |||||||||||||||| |||||| ||||||||||||||||||||| ||||
Db         531 CAGGTAGTAAAGAGAGCTCACTAAAACGCTAATTAGGCAAAAACAGGAGGTAAAGAAATA 472

Qy         718 GCCAATCATCTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAG 777
               ||||||||||||| |||||||||||||  |||| |||||| || | |||||||||| |||
Db         471 GCCAATCATCTATCGCCTGAGAGCACACTGGGAGGGACAATGA-TAGGATATAAACCCAG 413

Qy         778 GCATTCAAGCCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGT 837
               |||||||| |||||||  || |||| ||||| |||||||||| |||||||||||||||||
Db         412 GCATTCAAACCAGCAAGGGCTACCCTCTTTGAGTCCCCTCCCTTTGTATGGGAGCTCTGT 353

Qy         838 TTTCACTCTATTTCACTCTATTAAATCATGCAACTG--CACTCTTCTGGTCCGTGTTTTT 895
                         ||||||||||||||||| ||||||||  ||||||||||||| |||||| |
Db         352 ----------TTTCACTCTATTAAATCTTGCAACTGCACACTCTTCTGGTCAGTGTTTGT 303

Qy         896 TATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCCACCGTCACAGAC 955
               || | ||| |||||||||||   | ||| ||||||||||||||||||| || || |||||
Db         302 TACGACTCGAGCTGAGCTTTCACTTGCCGTCCACCACTGCTGTTTGCCGCCATCGCAGAC 243

Qy         956 CCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGC 1015
               |||| ||||||||||| ||||   ||||  |||| |||||  ||||| |||||||||||
Db         242 CCGCCGCTGACTTCCACCCCTCCAGATCTGGCAGGGTGTCTGCTGTGATCCTGATCCAGT 183

Qy        1016 GAGGTACCCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAA 1075
               |||| |||||||||| |||  |||  |||||||| |||| |||||||| |||| ||||||
Db.        182 GAGGCACCCATTGCCGCTCTGGATTGGGCTAAAGTCTTGTCATTGTTCTTGCACGGCTAA 123

Qy        1076 GTGCCTGGGTTTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCC 1135
               ||||| |||||  | ||||| || ||||||| | ||| ||||||||||  |||||||||
Db         122 GTGCCCGGGTTCATTCTAATCGAGCTGAACATTAGTCGCTGGGTTCCACAGTTCTCTTCT 63

Qy        1136 ATGACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCAT 1191
                 ||||||||||||||||||||||||||||   ||| |||||||||||||||||||
Db          62 GAGACCCACGGCTTCTAATAGAGCTATAACGAGCACTGCATGGCCCAAGATTCCAT 7


RESULT 29
AG076758/c
LOCUS       AG076758                 679 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-071A23.R, genomic survey sequence.
ACCESSION   AG076758
VERSION     AG076758.1  GI:16628560
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 679)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector     : pKS145
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1. .679
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-071A23.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             32.8%;  Score 435.6;  DB 9;  Length 679;
  Best Local Similarity   86.3%;  Pred. No. 1.9e-120;
  Matches  480;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy         133 AATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCTGGGGCCTCCTCAGCCAATG 192
               ||   ||||||||||| || |  |||| | | | |  |  ||||||||| |||| |||||
Db         599 AACTCAATGTAGAGCATAGTAGTTTCAGAGCGCCGGTCATTGGGGCCTCGTCAGTCAATG 540

Qy         193 GATGCCCTGGACTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTTTTACTCCTCTT 252
               ||||||||||  || ||||||||||| |||| |||||||| |||   | || ||||||||
Db         539 GATGCCCTGGGTTCGCCCCTTCTTAGAACCTGTAGCAGCTGTAAGGCTGTTGCTCCTCTT 480

Qy         253 TGGACCCTGTATCTTCAACTTCCTTGTTAAGTTTGTCTCTTCCAGAATTGAAGCTGTAAA 312
               ||| |||||||| ||| || |||||||| |||||||||||||||||||| | ||||||||
Db         479 TGGGCCCTGTATATTCGACCTCCTTGTTGAGTTTGTCTCTTCCAGAATTAAGGCTGTAAA 420

Qy         313 GCTACAAATAGTTCTTCAAATGGAACCCCAGATGCAGTCCATGACTAAAATCTACCGTGG 372
               ||||||||| |||||||||||| | |||||||||||||||||| |||||||||||| ||
Db         419 GCTACAAATGGTTCTTCAAATGAAGCCCCAGATGCAGTCCATGGCTAAAATCTACCATGT 360

Qy         373 ACCCCTGGACCGGCCTGCTAGACTATGCTCTGATGTTAATGACATTGAAGTCACCCCTCC 432
               |||||||| || ||||||||| |  ||||| ||||||||||||||  ||| |||||||||
Db         359 ACCCCTGGGCCAGCCTGCTAGCCCTTGCTCCGATGTTAATGACATCAAAGGCACCCCTCC 300

Qy         433 CGAGGAAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAG 492
                |||||||||||||||||||||||||||| |  | ||| ||||| ||||||||| |||||
Db         299 TGAGGAAATCTCAACTGCACAACCCCTACCATGCCCCATTTCAGCAGGAAGCAGGTAGAG 240

Qy         493 CAGTTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTTCCTGTTGAGAGGGTGGACTGAG 552
               | || ||||| ||||||||||| ||| |||||||||||||||||||||||| | ||||||
Db         239 CGGTCGTCAGTCAACCTCCCCAGCAGCACTTGGGTTTTCCTGTTGAGAGGGGGTACTGAG 180

Qy         553 AGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGGGAAGGT 612
               ||||||||||||||||||||||||| | ||||||||||||| ||||||| ||||||||||
Db         179 AGACAGGACTAGCTGGATTTCCTAGACCGACTAAGAATCCCTAAGCCTAGCTGGGAAGGT 120

Qy         613 GACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGA 672
               |||||| |||| ||||||||| ||||||||||||||||||||||||||||||||||||||
Db         119 GACCGCGTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGA 60

Qy         673 GCTCACTAAAATGCTA 688
               |||| | |  ||||||
Db          59 GCTCTCGAGCATGCTA 44


RESULT 30
AG102951
LOCUS       AG102951                 683 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-106G16.F, genomic survey sequence.
ACCESSION   AG102951
VERSION     AG102951.1  GI:16723468
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 683)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
```

```
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector     : pKS145
                R.Site 1   : SacI
                R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1..683
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-106G16.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             32.6%;  Score 433.6;  DB 9;  Length 683;
  Best Local Similarity   88.6%;  Pred. No. 7.8e-120;
  Matches  514;  Conservative    0;  Mismatches   61;  Indels    5;  Gaps    4;

Qy         547 ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAAGCCTANCTGG 606
               | ||||||| ||||||||||||||||||||||| ||||||||||||| ||||||| ||||
Db         108 ATTGAGAGAAAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGG 167

Qy         607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT 666
               ||||||||| ||||||| ||||||||||||||||||||||| ||||||||||| ||||||
Db         168 GAAGGTGACTGCATCCACCTTTAAACATGGGGCTTGCAACTGAGCTCACACCCAACCAAT 227

Qy         667 CAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCAT 726
               |||||||||||||||||||||||| |||||||||||||| |||||| || ||||||||||
Db         228 CAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGA-GTAAAGAAACAGCCAATCAT 286

Qy         727 CTATTGCCTGAGAGCACAGCGGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAAG 786
               |||||||||||||||||||||||| |||||||||| ||||||||||| ||||||||  ||
Db         287 CTATTGCCTGAGAGCACAGCGGGAGGGACAAGGATCGGGATATAAACCCAGGCATTTGAG 346

Qy         787 CCAGCAACAGCAACCCCCTTTGGGTCCCCTCCCATTGTATGGGAGCTCTGTTTTCACTCT 846
               || ||||| |||||||||||||||||||||||  |||||||||| |||||||||||||||
Db         347 CCGGCAACGGCAACCCCCTTTGGGTCCCCTCC--TTGTATGGGAACTCTGTTTTCACTCT 404

Qy         847 ATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTTATGGCTCAAG 906
               |||||||||||||||||| ||||||||  |||||||||||||||||| | ||||||||||
Db         405 ATTTCACTCTATTAAATCTTGCAACTGTGCTCTTCTGGTCCGTGTTTGTAATGGCTCAAG 464

Qy         907 CTGAGCTTTTGTTCGCCATCCACCACTGCTG-TTTGCCACCGTCACAGACCCGCTGCTGA 965
               |||||||||   | ||  ||||||||||||| ||||   ||  | |  ||||||| | |||||
Db         465 CTGAGCTTTCACTTGCTGTCCACCACTGCTGTTTTGTTGCCATTGCAGACCCACCGCTGA 524

Qy         966 CTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGTACCCA 1025
               |||||||||||  ||||||||||||||||   ||| ||||||||||||| |||   ||||
Db         525 CTTCCATCCCTCCGGATCCAGCAGAGTGTGGGCTGGGCTCCTGATCCAGTGAGACGCCCA 584

Qy        1026 TTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGGT 1085
               |||||  | |||||  |||||||||||||||||||||||||||||||||||||||||| |
Db         585 TTGCC-GTTCCGATTGGGCTAAAGGCTTGCCATTGTTCCTGCATGGCTAAGTGCCTGGAT 643

Qy        1086 TTGTCCTAATAGAACTGAACACTGGTCACTGGGTTCCATG 1125
               |  ||||||| || ||||||||| ||||||| |||||| |
Db         644 TCATCCTAATCGAGCTGAACACTAGTCACTGTGTTCCAGG 683


Search completed: January 22, 2005, 19:31:51
Job time : 3968.84 secs
```